LIGAND BINDING DOMAINS OF NUCLEAR HORMONE RECEPTORS

This invention relates to a protein, termed CAA05410.2 and to functional equivalents of this protein, herein identified as containing novel Nuclear Hormone Receptor Ligand Binding Domains and more preferably as nuclear receptors and to the use of these proteins and nucleic acid sequences from the encoding genes in the diagnosis, prevention and treatment of disease. CAA05410.2 has been identified as containing both a Nuclear Hormone Receptor Ligand Binding Domain and a Nuclear Hormone Receptor DNA Binding Domain.

All publications, patents and patent applications cited herein are incorporated in full by reference.

BACKGROUND

The process of drug discovery is presently undergoing a fundamental revolution as the era of functional genomics comes of age. The term "functional genomics" applies to an approach utilising bioinformatics tools to ascribe function to protein sequences of interest. Such tools are becoming increasingly necessary as the speed of generation of sequence data is rapidly outpacing the ability of research laboratories to assign functions to these protein sequences.

As bioinformatics tools increase in potency and in accuracy, these tools are rapidly replacing the conventional techniques of biochemical characterisation. Indeed, the advanced bioinformatics tools used in identifying the present invention are now capable of outputting results in which a high degree of confidence can be placed.

Various institutions and commercial organisations are examining sequence data as they become available and significant discoveries are being made on an on-going basis. However, there remains a continuing need to identify and characterise further genes and the polypeptides that they encode, as targets for research and for drug discovery.

Recently, a remarkable tool for the evaluation of sequences of unknown function has been developed by the Applicant for the present invention. This tool is a database system, termed the Biopendium search database, that is the subject of co-pending International Patent Application No. PCT/GB01/01105. This database system consists of an integrated data resource created using proprietary technology and containing information generated

from an all-by-all comparison of all available protein or nucleic acid sequences.

The aim behind the integration of these sequence data from separate data resources is to combine as much data as possible, relating both to the sequences themselves and to information relevant to each sequence, into one integrated resource. All the available data relating to each sequence, including data on the three-dimensional structure of the encoded protein, if this is available, are integrated together to make best use of the information that is known about each sequence and thus to allow the most educated predictions to be made from comparisons of these sequences. The annotation that is generated in the database and which accompanies each sequence entry imparts a biologically relevant context to the sequence information.

This data resource has made possible the accurate prediction of protein function from sequence alone. Using conventional technology, this is only possible for proteins that exhibit a high degree of sequence identity (above about 20%-30% identity) to other proteins in the same functional family. Accurate predictions are not possible for proteins that exhibit a very low degree of sequence homology to other related proteins of known function.

In the present case, a protein whose sequence is recorded in a publicly available database as CAA05410.2 (NCBI Genebank nucleotide accession number AJ002425.2 and a Genebank protein accession number CAA05410.2), is implicated as containing a novel Nuclear Hormone Receptor Ligand Binding Domain.

Introduction to Nuclear Hormone Receptor Ligand Binding Domains

The Nuclear Hormone Receptor superfamily (see Table 1) encodes structurally related proteins that regulate the transcription of target genes. These proteins include receptors for steroid and thyroid hormones, vitamins, and other proteins for which no ligands have been found. To be classified as a "Nuclear Hormone Receptor" a protein must possess at least one of two key domains; a C4-type zinc finger DNA-Binding Domain (DBD) or a Ligand Binding Domain (LBD). The DBD is required for binding DNA in the vicinity of target genes, and the LBD is required for steroid-like ligand responsiveness. It is the Ligand Binding Domain of Nuclear Hormone Receptors which is the binding site for pharmacological agents such as Tamoxifen.

Many Nuclear Hormone Receptors possess both a DBD and an LBD, and a well-known example of this is Estrogen receptor alpha, which possesses both a DBD and an LBD. Nuclear Hormone Receptors which possess both a DBD and an LBD can be referred to as "Classical Nuclear Hormone Receptors". There are also members of the Nuclear Hormone Receptor family which possess a DBD but lack an LBD; for example the proteins Knirps (SWISS-PROT code P10734) and ODR7 (SWISS-PROT code P41933) possess DBDs, but both lack LBDs. Implicit in the existence of proteins such as Knirps and ODR7 is the fact that possession of a DBD does not mean that a LBD will be concomitantly present.

There are also members of the Nuclear Hormone Receptor family that possess an LBD but lack a DBD; for example the protein "Short Heterodimer Partner", SHP (SWISS-PROT code Q15466) possesses an LBD but lacks a DBD.

A further refinement in the classification of Nuclear Hormone Receptors is to classify on the basis of possession of an LBD. Nuclear Hormone Receptors which possess an LBD can be sub-classified as "Nuclear Hormone Receptor Ligand Binding Domain" family members. Thus Estrogen receptor alpha and SHP are "Nuclear Hormone Receptor Ligand Binding Domain" family members whereas Knirps and ODR7 are excluded. Similarly, Nuclear Hormone Receptors can also be sub-classified on the basis of possession of a DBD. Nuclear Hormone Receptors which possess a C4-type zinc finger DBD can be sub-classified as "Nuclear Hormone Receptor DNA Binding Domain" family members. Thus Estrogen Receptor alpha, Knirps and ODR7 are "Nuclear Hormone Receptor DNA Binding Domain" family members whereas SHP is excluded.

The DBD directs the protein to bind specific DNA sequences in the vicinity of target genes.

The Ligand Binding Domain (LBD) binds and responds to the cognate hormone. Ligand binding to the LBD triggers a conformational change which expels a bound "Nuclear Receptor Co-Repressor". The site previously occupied by the Co-Repressor is then free to recruit a "Nuclear Receptor Co-Activator". This Ligand-triggered swap of a Co-Repressor for a Co-Activator is the mechanism by which Ligand binding leads to the transcriptional activation of target genes. The LBD is the binding site for all Nuclear Hormone Receptor targeted drugs to date and it is thus desirable to identify novel Ligand

Binding Domains since these will be attractive drug targets. The LBD also directs dimerisation with other LBDs. For example, the Estrogen receptor alpha ligand binding domain can homodimerise with itself, or heterodimerise with the Estrogen receptor beta ligand binding domain. Ligand Binding Domains share low sequence identity (~15%) but have very similar structures and so present ideal targets for a structure-based relationship tool such as Inpharmatica Genome Threader™.

Table 1: Nuclear Hormone Receptor Superfamily

Class ID	Nuclear Hormone Receptor Name	Species	Accession
NR1 group			
NR1A1	Thyroid Hormone Receptor alpha	HUMAN	M24748
NR1A2	Thyroid Hormone Receptor beta	HUMAN	X04707
NR1A3	Thyroid Hormone Receptor Ciona	CIONA	AF077403
NR1B1	Retinoic Acid Receptor alpha	HUMAN	X06538
NR1B2	Retinoic Acid Receptor beta	HUMAN	Y00291
NR1B3	Retinoic Acid Receptor gamma	HUMAN	M57707
NR1B4	Retinoic Acid Receptor Polyandrocarpa	POLYANDROCARPA	D86615
NR1C1	Peroxisome Proliferator Activated Receptor alpha	HUMAN	L02932
NR1C2	Peroxisome Proliferator Activated Receptor beta	HUMAN	L07592
NR1C3	Peroxisome Proliferator Activated Receptor gamma	HUMAN	L40904
NR1D1	Rev-erbA	HUMAN	M24898
NR1D2	Rev-erbB	HUMAN	L31785
NR1D3	E75	FLY	X51548
NR1E1	E78	FLY	U01087
NR1F1	RAR-related Orphan Receptor alpha	HUMAN	U04897
NR1F2	RAR-related Orphan Receptor beta	HUMAN	Y08639
NR1F3	RAR-related Orphan Receptor gamma	HUMAN	U16997
NR1F4	DHR3	FLY	M90806
NR1G1	CNR14	WORM	U13074
NR1H1	Ecdysone Receptor	FLY	M74078
NR1H2	Liver X Receptor beta	HUMAN	U07132
NR1H3	Liver X Receptor alpha	HUMAN	U22662
NR1H4	Farnesoid X Receptor	HUMAN	U68233
NR1I1	Vitamin D Receptor	HUMAN	J03258
NR1I2	Pregnane X Receptor	HUMAN	AF061056
NR1I3	Constitutive Androstane Receptor alpha	HUMAN	Z30425
NR1I4	Constitutive Androstane Receptor beta	MOUSE	AF009327
NR1J1	DHR96	FLY	U36792
NR1K1	NHR1	WORM	U19360
NR2 group			
NR2A1	Hepatocyte Nuclear Factor 4 alpha	HUMAN	X76930
NR2A2	Hepatocyte Nuclear Factor 4 beta	XENOPUS	Z49827
NR2A3	Hepatocyte Nuclear Factor 4 gamma	HUMAN	Z49826
NR2A4	Drosophila Hepatocyte Nuclear Factor 4	FLY	U70874
NR2B1	Retinoid X Receptor alpha	HUMAN	X52773

NR2B2	Retinoid X Receptor beta	HUMAN	M84820
NR2B3	Retinoid X Receptor gamma	HUMAN	U38480
NR2B4	Ultraspiracle	FLY	X53417
NR2C1	TR2	HUMAN	M29960
NR2C2	TR4	HUMAN	L27586
NR2D1	SpSHR2	SEAURCHIN	U38281
NR2E1	TLX	HUMAN	Y13276
NR2E2	Tailless	FLY	AF019362
NR2E3	Photoreceptor-specific Nuclear Receptor	HUMAN	AF121129
NR2E4	Dissatisfaction	FLY	AF106677
NR2E5	FAX-1	WORM	AF176087
NR2F1	COUP-TFI	HUMAN	X12795
NR2F2	COUP-TFII	HUMAN	M64497
NR2F3	Seven-up	FLY	M28863
NR2F4	Xenopus COUP-TFIII	XENOPUS	X63092
NR2F5	Zebrafish COUP-TFIII	ZEBRAFISH	X70300
NR2F6	EAR2	HUMAN	X12794
NR3 group			
NR3A1	Estrogen Receptor alpha	HUMAN	P03372
NR3A2	Estrogen Receptor beta	HUMAN	AB006590
NR3B1	Estrogen Receptor Related alpha	HUMAN	X51416
NR3B2	Estrogen Receptor Related beta	HUMAN	AF094517
NR3B3	Estrogen Receptor Related gamma	HUMAN	AF058291
NR3C1	Glucocorticoid Receptor	HUMAN	X03225
NR3C2	Mineralocorticoid Receptor	HUMAN	M16801
NR3C3	Progesterone Receptor	HUMAN	M15716
NR3C4	Androgen Receptor	HUMAN	M20132
NR4 group			
NR4A1	NGFI-Balpha	HUMAN	L13740
NR4A2	NGFI-Bbeta	HUMAN	X75918
NR4A3	NGFI-Bgamma	HUMAN	D78579
NR4A4	DHR38	FLY	X89246
NR5 group			
NR5A1	FTZ-F1	HUMAN	U76388
NR5A2	FTF	HUMAN	U93553
NR5A3	Drosophila FTZ-F1	FLY	M98397
NR5A4	Zebrafish FTZ-F1	ZEBRAFISH	AF198086
NR5B1	FTZ-F1B	FLY	L06423
NR6 group			
NR6A1	Germ Cell Nuclear Factor	HUMAN	U64876
NR6A2	GCNF Related Factor	TENEBRIO	AF124981
NR0 group (have only one characteristic domain)			
NR0A1 sub-group (have only DBD no LBD)			
NR0A1	Knirps	FLY	X13331
NR0A2	Knirps-related	FLY	X14153
NR0A3	Embryonic gonad	FLY	X16631
NR0A4	ODR7	WORM	U16708
NR0B sub-group (have only LBD no DBD)			
NR0B1	DAX1	HUMAN	S74720
NR0B2	Short Heterodimer Partner SHP	HUMAN	L76571

Table data taken from Laudet and Gronemeyer "The Nuclear Receptor Facts Book", Academic Press. Class ID refers to a classification code for each member, and accession refers to NCBI GenBank nucleotide accession code.

II. Nuclear Hormone Receptor Ligand Binding Domain Family and Disease

Nuclear Hormone Receptor Ligand Binding Domain family members have been shown to play a role in diverse physiological functions, many of which can play a role in disease processes (see Table 2).

Table 2. Nuclear Hormone Receptors and disease

Nuclear Hormone Receptor	Disease
Androgen Receptor	Androgen Insensitivity Syndrome (Lubahn *et al.* 1989 Proc. Natl. Acad. Sci. USA 86, 9534-9538).
	Reifenstein syndrome (Wooster *et al.* 1992 Nat. Genet. 2, 132-134).
	X-linked recessive spinal and bulbar muscular atrophy (MacLean *et al.* 1995 Mol. Cell. Endocrinol. 112,133-141).
	Male breast cancer ((Wooster *et al.* 1992 Nat. Genet. 2, 132-134).
Glucocorticoid Receptor	Nelson's syndrome (Karl *et al.* 1996 J. Clin. Endocrinol. Metab. 81, 124-129).
	Glucocorticoid resistant acute T-cell leukemia (Hala *et al.* 1996 Int. J. Cancer 68, 663-668).
Mineralocorticoid Receptor	Pseudohypoaldosteronism (Chung *et al.* 1995 J. Clin. Endocrinol. Metab. 80, 3341-3345).

Estrogen Receptor alpha and beta	ER alpha expression is elevated in a subset of human breast cancers. The application of Tamoxifen is the major therapy to prevent breast tumour progression. (Petrangeli *et al.* 1994 J. Steroid Biochem. Mol. Biol. 49, 327-331). Estrogen Receptors are involved in cancer, particular cancers originating from estrogen-responsive tissues, including breast (as mentioned above), uterus and prostate, myeloproliferative disorders, such as leukemia, hypertension, hypotension, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, wound healing, scarring, obesity, dermatological disorders including cellulite, estrogen-mediated hair characteristics, central nervous system disorders, Alzheimer's disease, cognition enhancement, learning and memory enhancement, immunomodulation, and osteoporosis (taken from Laudet and Gronemeyer, "The Nuclear Receptor Facts Book" Academic Press).
Vitamin D3 Receptor	Mutations in the Vitamin D3 receptor produce a hereditary disorder similar in phenotype to Vitamin D3 deficiency (Rickets) (Hughes *et al.* 1988 Science 242, 1702-1725).
Retinoic Acid Receptor alpha	Acute Myeloid Leukemia (Lavau and Dejean 1994 Leukemia 8, 9-15).
Thyroid Hormone Receptor beta	"Generalised Resistance to Thyroid Hormones" (GRTH) (Refetoff 1994 Thyroid 4, 345-349).
DAX1	X-linked Adrenal Hypoplasia Congenita (AHC) and Hypogonadism (Ito *et al.* 1997 Mol. Cell. Biol. 17, 1476-1483).

Alteration of Nuclear Hormone Receptor family members by ligands which bind to their LBD thus provides a means to alter the disease phenotype. There is thus a great need for the identification of novel Nuclear Hormone Receptors, as these proteins may play a role in the diseases identified above, as well as in other disease states. The identification of novel Nuclear Hormone Receptors is thus highly relevant for the treatment and diagnosis of disease, particularly those identified in Table 2.

THE INVENTION

The invention is based on the discovery that a region of the CAA05410.2 protein functions as a Nuclear Hormone Receptor. CAA05410.2 has been identified as containing both a Nuclear Hormone Receptor Ligand Binding Domain and a Nuclear Hormone

Receptor DNA Binding Domain.

For the CAA05410.2 protein, it has been found that a region including residues 394-604 of this protein sequence adopts an equivalent fold to residues 15 to 247 (equivalent to residues 320 to 552 in the full-length numbering scheme) of the Human Estrogen Receptor alpha Ligand Binding Domain (PDB code 3ERT:A). Human Estrogen Receptor alpha Ligand Binding Domain is known to function as a Nuclear Hormone Receptor Ligand Binding Domain. This relationship is not just to the Human Estrogen Receptor alpha Ligand Binding Domain, but rather to the Nuclear Hormone Receptor Ligand Binding Domain family as a whole.

The discovery of sharing an equivalent fold allows the functional annotation of this region of CAA05410.2, and therefore proteins that include this region, as possessing Nuclear Hormone Receptor Ligand Binding Domain activity.

In addition, results presented herein clearly indicate that the transcript for LBDG11 (also referred to herein as NHR11 or NR11), a human paralogue of the CAA05410.2 protein, is present at detectable levels in a variety of human tissues and cell lines. This confirms the relevance of the LBDG11 polypeptides as important targets for further biochemical characterisation. The particular tissues and cell lines identified herein as expressing LBDG11 represent ideal targets for further studies of LBDG11 function *in vivo*. Such studies may, for example, make use of the ligands identified using the assays and screening methods disclosed herein to investigate the effects of inducing or inhibiting LBDG11 function. In addition, the cloning of the full-length LBDG11 polypeptide allows for high-level expression, purification and characterisation of the LBDG11 polypeptides of the invention. For example, the cloning, purification and partial characterisation of the LBD of LBDG11 is described herein. Although the Applicant does not wish to be bound by this theory, the LBDG11 polypeptide may be involved in DNA repair and may interact with BRCA polypeptides.

Accordingly, agonists and antagonists of LBDG11 are likely to be of great value in the treatment of diseases in which Nuclear Hormone Receptors are implicated, especially tumours, such as prostate cancers, breast cancer and cervical cancer. As described above, agonists and antagonists of the LBDG11 polypeptides can be readily identified using the assays and screening methods disclosed herein. Once identified, the effect of agonists and

antagonists on diseased cell lines and tissue types may then be investigated using the methods disclosed herein or known to those of skill in the art. It is likely that certain agonists or antagonists identified using the assays and methods disclosed herein will be useful in the prophylaxis or treatment of diseases associated with LBDG11.

For example, the present invention allows the development of molecular diagnostic tools, such as monoclonal antibodies, for the detection of LBDG11 *in vivo*, that preferably target the ligand binding domain or DNA binding domain specifically. It is the teaching of the invention, that various related polypeptides function as Nuclear Hormone Receptors, that allows the skilled reader to use this knowledge to generate bespoke compounds that bind to areas of interest and biochemical importance in the polypeptide. Furthermore, the identification of patients affected by particular cancer conditions using such diagnostic methods will enable specific therapeutic approaches for individual patients to be selected.

For example, the present invention allows the design of specific therapies for the treatment of the specific cancer conditions identified herein; such therapies may, for example, target LBDG11 expression or function in the relevant tumour cells.

In a first aspect, the invention provides a polypeptide, which polypeptide:

(i) comprises the amino acid sequence as recited in SEQ ID NO:2;

(ii) is a fragment thereof having activity as a Nuclear Hormone Receptor Ligand Binding Domain and/or a Nuclear Hormone Receptor DNA Binding Domain and/or a Nuclear Hormone Receptor or having an antigenic determinant in common with the polypeptides of (i); or

(iii) is a functional equivalent of (i) or (ii).

Preferably, a polypeptide according to the first aspect of the invention consists of the amino acid sequence as recited in SEQ ID NO:2 or is a fragment or functional equivalent thereof.

The polypeptide having the sequence recited in SEQ ID NO:2 is referred to hereafter as "the LBDG5 polypeptide".

Preferred functionally-equivalent homologues of the LBDG5 polypeptide are the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide and the

LBDG13 polypeptide referred to above. The LBDG10 polypeptide has 83% sequence identity with the LBDG5 polypeptide. A polypeptide consisting of the amino acid sequence as recited in SEQ ID NO:6 (the LBDG11 polypeptide) or a fragment or functional equivalent thereof, is a particularly preferred functional equivalent of the LBDG5 polypeptide.

In the present case, preferred active fragments of the LBDG5 polypeptide are those that include the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region. As defined herein, the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 394 and residue 604 of the LBDG5 polypeptide sequence.

Preferred active fragments of the LBDG10 polypeptide are those that include the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region. As discussed above, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 413 and residue 603 of the LBDG10 polypeptide sequence.

Preferred active fragments of the LBDG11 polypeptide are those that include the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region. As discussed above, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 118 and residue 319 of the LBDG11 polypeptide sequence.

Preferred active fragments of the LBDG12 polypeptide are those that include the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region. As discussed above, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 118 and residue 318 of the LBDG12 polypeptide sequence.

Preferred active fragments of the LBDG13 polypeptide are those that include the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region. As discussed above, the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 117 and residue 317 of the LBDG13 polypeptide sequence.

This aspect of the invention also includes fusion proteins that incorporate polypeptide fragments and variants of these polypeptide fragments as defined above, provided that said fusion proteins possess activity as a Nuclear Hormone Receptor Ligand Binding Domain and/or a Nuclear Hormone Receptor DNA Binding Domain and/or a Nuclear Hormone Receptor.

When referring to polypeptides with 'activity as a Nuclear Hormone Receptor Ligand Binding Domain' herein, we refer to polypeptides that comprise amino acid sequence or structural features that can be identified as conserved features present in the polypeptides disclosed herein and other known Nuclear Hormone Receptor Ligand Binding Domains, and that also retain their ligand-binding specificity such that LBD function is not substantially reduced in comparison to the naturally-occurring LBD. For example, assays to test ligand binding and its effect on LBDG11-directed transcription are described in the Examples.

When referring to polypeptides with 'activity as a Nuclear Hormone Receptor DNA Binding Domain' herein, we refer to polypeptides that comprise amino acid sequence or structural features that can be identified as conserved features present in both the polypeptides of the invention and other known Nuclear Hormone Receptor DNA Binding Domains, and that also retain their DNA-binding specificity such that DBD function is not substantially reduced in comparison to the naturally-occurring DBD.

When referring to polypeptides with 'activity as a Nuclear Hormone Receptor' herein, we refer to polypeptides that comprise amino acid sequence or structural features that can be identified as conserved features present in the polypeptides of the invention and in other known Nuclear Hormone Receptors, and that also retain their ligand-binding and DNA-binding specificity such that NHR function is not substantially reduced in comparison to the naturally-occurring polypeptide. For example, assays to evaluate the ability of a polypeptide for activity as a Nuclear Hormone Receptor, mediating effects on transcription through ligand binding and subsequent binding of the DNA binding domain to a DNA target are described in the Examples.

In a second aspect, the invention provides a purified nucleic acid molecule that encodes a polypeptide of the first aspect of the invention, or a fragment thereof. Preferably, the purified nucleic acid molecule has the nucleic acid sequence as recited in SEQ ID NO:1

(encoding the LBDG5 polypeptide), SEQ ID NO:3 (encoding the LBDG10 polypeptide), SEQ ID NO:5 (encoding the LBDG11 polypeptide), SEQ ID NO:7 (encoding the LBDG12 polypeptide), SEQ ID NO:9 (encoding the LBDG13 polypeptide), or is a redundant equivalent or fragment of this sequence. A further preferred nucleic acid molecule is one that encodes a polypeptide fragment according to part ii) above, preferably a polypeptide fragment that includes the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, or that encodes a variant of this fragment as this term is defined above.

In a third aspect, the invention provides a purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule of the second aspect of the invention.

In a fourth aspect, the invention provides a vector, such as an expression vector, that contains a nucleic acid molecule of the second or third aspect of the invention.

In a fifth aspect, the invention provides a host cell transformed with a vector of the fourth aspect of the invention.

In a sixth aspect, the invention provides a ligand which binds specifically to, and which either inhibits or activates the Nuclear Hormone Receptor Ligand Binding Domain and/or a Nuclear Hormone Receptor DNA Binding Domain and/or a Nuclear Hormone Receptor activity of, a polypeptide of the first aspect of the invention. Preferably, the ligand inhibits the function of a polypeptide of the first aspect of the invention, namely, that of a nuclear hormone receptor protein.

Ligands to a polypeptide according to the invention may come in various forms, including natural or modified substrates, enzymes, receptors, small organic molecules such as small natural or synthetic organic molecules of up to 2000Da, preferably 800Da or less, peptidomimetics, inorganic molecules, peptides, polypeptides, antibodies, structural or functional mimetics of the aforementioned.

In a seventh aspect, the invention provides a compound that is effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

A compound of the seventh aspect of the invention may either increase (agonise) or decrease (antagonise) the level of expression of the gene or the activity of the polypeptide. Importantly, the identification of the function of the region defined herein as the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG10 polypeptide, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG11 polypeptide, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG12 polypeptide, and the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG13 polypeptide respectively, allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of diseases in which Nuclear Hormone Receptor Ligand Binding Domains are implicated. Ligands and compounds according to the sixth and seventh aspects of the invention may be identified using such methods. These methods are included as aspects of the present invention.

For example, such ligands may bind specifically to, and inhibit or activate the activity of a Nuclear Hormone Receptor Ligand Binding Domain of the present invention by binding to one or more residues within the LBD, or to one or more residues outside the LBD. As noted above, all known drugs that affect Nuclear Hormone Receptors bind to residues within the LBD. Accordingly, it is preferred that ligands of the invention which inhibit the Nuclear Hormone Receptor Ligand Binding Domain activity of a polypeptide of the invention bind to residues within the LBD.

For example, such ligands may bind specifically to, and inhibit the activity of a Nuclear Hormone Receptor DNA Binding Domain of the present invention by binding to one or more residues within the LBD, or to one or more residues outside the LBD. It is preferred that ligands of the invention which inhibit the Nuclear Hormone Receptor DNA Binding Domain activity of a polypeptide of the invention bind to residues within the DBD.

For example, such ligand may bind specifically to, and inhibit the activity of a Nuclear Hormone Receptor of the present invention by binding to either the LBD or the DBD, such that the NHR is unable to function normally.

In a seventh aspect, the invention provides a compound that is effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the

invention or to regulate the activity of a polypeptide of the first aspect of the invention.

Such compounds may be identified using the assays and screening methods dislcosed herein.

A compound of the seventh aspect of the invention may either increase (agonise) or decrease (antagonise) the level of expression of the gene or the activity of the polypeptide. Importantly, the identification of the function of the regions defined herein as the Nuclear Hormone Receptor Ligand Binding Domain regions of each polypeptide, allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of diseases in which Nuclear Hormone Receptors are implicated. Examples of suitable assays and screening methods are provided herein.

In an eighth aspect, the invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, for use in therapy or diagnosis of diseases in which Nuclear Hormone Receptors are implicated. Preferably the disease is proliferative in nature such as inflammation or cancer. Examples of diseases in which Nuclear Hormone Receptors are implicated include cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, uterus, prostate, cervical, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposis' sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, hypotension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, Parkinson's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, cognition enhancement, learning and memory enhancement, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, thyroid hormone

resistance syndrome, hypercalcemia, hypocalcaemia, hypercholestrolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, blood disorders including hemophilia, dermatological disorders, including, cellulite, acne, eczema, psoriasis and wound healing, scarring, negative effects of aging, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, hormone replacement therapies, steroid hormone-like mediated hair characteristics, immunomodulation, AIDS, vision disorders, glucocorticoid resistance, mineralocorticoid resistance, androgen resistance, pseudohypoaldosteronism, spinal/bulbar muscular atrophy, extraskeletal myxoid chrondrosarcomas, adrenal insufficiency, sexual reversal, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions.

The prediction that the LBDG5 polypeptide and its functional equivalent, the LBDG10 polypeptide form heterodimers with polypeptides that contain Nuclear Hormone Receptor Ligand Binding Domains, particularly ERa and ERß, also makes this polypeptide, the LBDG10 polypeptide and compounds of the nature described above that bind it of potential utility in the diagnosis and treatment of diseases in which ERa and ERß are implicated. Such diseases include cancer, particular cancers originating from estrogen-responsive tissues, including breast, uterus, cervix and prostate, myeloproliferative disorders, such as leukemia, hypertension, hypotension, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, wound healing, scarring, obesity, dermatological disorders including cellulite, estrogen-mediated hair characteristics, central nervous system disorders, Alzheimer's disease, cognition enhancement, learning and memory enhancement, immunomodulation, osteoporosis, among other pathological conditions.

LBDG11 has been shown to react with terminal deoxynucleotidyl transferase (TdT), a DNA-independent DNA polymerase that contributes to antigen-receptor diversity. The DNA polymerase mu is 42% identical to TdT so it is therefore likely that LBDG11 interacts with polymerase mu (which is involved in DNA repair, Ruiz *et al.* 2001 Philos. Trans. R. Soc. Lond. B. Biol. Sci. 356(1405):99-109) and is involved in its regulation. It can therefore be postulated that LBDG11 is involved in DNA repair and regulation.

In a ninth aspect, the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide of

the first aspect of the invention or the activity of a polypeptide of the first aspect of the invention in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease. Such a method will preferably be carried out *in vitro*. Similar methods may be used for monitoring the therapeutic treatment of disease in a patient, wherein altering the level of expression or activity of a polypeptide or nucleic acid molecule over the period of time towards a control level is indicative of regression of disease.

A preferred method for detecting polypeptides of the first aspect of the invention comprises the steps of: (a) contacting a ligand, such as an antibody, of the sixth aspect of the invention with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

A number of different such methods according to the ninth aspect of the invention exist, as the skilled reader will be aware, such as methods of nucleic acid hybridization with short probes, point mutation analysis, polymerase chain reaction (PCR) amplification and methods using antibodies to detect aberrant protein levels. Similar methods may be used on a short or long-term basis to allow therapeutic treatment of a disease to be monitored in a patient. The invention also provides kits that are useful in these methods for diagnosing disease.

A preferred method for detecting polypeptides of the first aspect of the invention comprises the steps of: (a) contacting a ligand of the sixth aspect of the invention with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex. In such methods the ligand of the sixth aspect of the invention may be any suitable ligand, such as an antibody, hormone or nucleic acid sequence.

Preferably, the disease diagnosed by a method of the ninth aspect of the invention is a disease in which Nuclear Hormone Receptors are implicated, as described above.

In a tenth aspect, the invention provides for the use of a polypeptide of the first aspect of the invention, or a fragment thereof, as a Nuclear Hormone Receptor Ligand Binding Domain and/or a Nuclear Hormone Receptor DNA Binding Domain and/or a Nuclear Hormone Receptor. Suitable uses of a polypeptide of the invention as a Nuclear Hormone

Receptor Ligand Binding Domain and/or a Nuclear Hormone Receptor DNA Binding Domain and/or a Nuclear Hormone Receptor include use for the hormone-mediated directed regulation of gene transcription. Other suitable uses of the polypeptides of the invention include use in methods for the identification of ligands, which ligands will be of use in the modulation of disease processes in which these polypeptides are implicated. For example, ligands for the LBDG11 nuclear hormone receptor ligand binding domain may be identified using the assays and screening methods disclosed herein. Such ligands may then be used in combination with the LBDG11 polypeptides to effect ligand-regulated expression of target genes.

The invention also provides for the use of a nucleic acid molecule according to the second or third aspects of the invention to express a protein that possesses activity as a Nuclear Hormone Receptor Ligand Binding Domain and/or a Nuclear Hormone Receptor DNA Binding Domain and/or a Nuclear Hormone Receptor. Such nucleic acid molecules are of utility in the production of the polypeptides of the invention, which polypeptides are useful in a variety of situations, as described above.

The invention also provides a method for effecting Nuclear Hormone Receptor Ligand Binding Domain and/or Nuclear Hormone Receptor DNA Binding Domain and/or Nuclear Hormone Receptor activity, said method utilising a polypeptide of the first aspect of the invention, or a fragment thereof.

In an eleventh aspect, the invention provides a pharmaceutical composition comprising a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell according to the fifth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, in conjunction with a pharmaceutically-acceptable carrier.

In a twelfth aspect, the present invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell according to the fifth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, for use in the manufacture of a medicament for the diagnosis or treatment of a disease in which Nuclear Hormone Receptors are implicated,

examples of which are given above.

In a thirteenth aspect, the invention provides a method of treating a disease in a patient comprising administering to the patient a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, or a pharmaceutical composition of the eleventh aspect of the invention.

For diseases in which the expression of a natural gene encoding a polypeptide of the first aspect of the invention, or in which the activity of a polypeptide of the first aspect of the invention, is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, host cell, ligand or compound administered to the patient should be an agonist. Conversely, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, host cell, ligand or compound administered to the patient should be an antagonist. Examples of such antagonists include antisense nucleic acid molecules, ribozymes and ligands, such as antibodies.

Preferably, the disease is a disease in which Nuclear Hormone Receptors are implicated, as described above.

In a fourteenth aspect, the invention provides transgenic or knockout non-human animals that have been transformed to express higher, lower or absent levels of a polypeptide of the first aspect of the invention. Such transgenic animals are very useful models for the study of disease and may also be used in screening regimes for the identification of compounds that are effective in the treatment or diagnosis of such a disease.

Preferably, the disease is a disease in which Nuclear Hormone Receptors are implicated, as described above.

A summary of standard techniques and procedures which may be employed in order to utilise the invention is given below. It will be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors and reagents described. It is

also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and it is not intended that this terminology should limit the scope of the present invention. The extent of the invention is limited only by the terms of the appended claims.

Standard abbreviations for nucleotides and amino acids are used in this specification.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA technology and immunology, which are within the skill of those working in the art.

Such techniques are explained fully in the literature. Examples of particularly suitable texts for consultation include the following: Sambrook Molecular Cloning; A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and II (D.N. Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 & 155; Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds. 1987, Academic Press, London); Scopes, (1987) Protein Purification: Principles and Practice, Second Edition (Springer Verlag, N.Y.); and Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell eds. 1986).

As used herein, the term "polypeptide" includes any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.* peptide isosteres. This term refers both to short chains (peptides and oligopeptides) and to longer chains (proteins).

The polypeptide of the present invention may be in the form of a mature protein or may be a pre-, pro- or prepro- protein that can be activated by cleavage of the pre-, pro- or prepro- portion to produce an active mature polypeptide. In such polypeptides, the pre-, pro- or prepro- sequence may be a leader or secretory sequence or may be a sequence that

is employed for purification of the mature polypeptide sequence.

The polypeptide of the first aspect of the invention may form part of a fusion protein. For example, it is often advantageous to include one or more additional amino acid sequences which may contain secretory or leader sequences, pro-sequences, sequences which aid in purification, or sequences that confer higher protein stability, for example during recombinant production. Alternatively or additionally, the mature polypeptide may be fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

Polypeptides may contain amino acids other than the 20 gene-encoded amino acids, modified either by natural processes, such as by post-translational processing or by chemical modification techniques which are well known in the art. Among the known modifications which may commonly be present in polypeptides of the present invention are glycosylation, lipid attachment, sulphation, gamma-carboxylation, for instance of glutamic acid residues, hydroxylation and ADP-ribosylation. Other potential modifications include acetylation, acylation, amidation, covalent attachment of flavin, covalent attachment of a haeme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulphide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, GPI anchor formation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl terminus in a polypeptide, or both, by a covalent modification is common in naturally-occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention.

The modifications that occur in a polypeptide often will be a function of how the polypeptide is made. For polypeptides that are made recombinantly, the nature and extent of the modifications in large part will be determined by the post-translational

modification capacity of the particular host cell and the modification signals that are present in the amino acid sequence of the polypeptide in question. For instance, glycosylation patterns vary between different types of host cell.

The polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally-occurring polypeptides (for example purified from cell culture), recombinantly-produced polypeptides (including fusion proteins), synthetically-produced polypeptides or polypeptides that are produced by a combination of these methods.

The functionally-equivalent polypeptides of the first aspect of the invention may be polypeptides that are homologous to the LBDG5, LBDG10, LBDG11, LBDG12 and/or LBDG13 polypeptidea. Two polypeptides are said to be "homologous", as the term is used herein, if the sequence of one of the polypeptides has a high enough degree of identity or similarity to the sequence of the other polypeptide. "Identity" indicates that at any particular position in the aligned sequences, the amino acid residue is identical between the sequences. "Similarity" indicates that, at any particular position in the aligned sequences, the amino acid residue is of a similar type between the sequences. Degrees of identity and similarity can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991).

Homologous polypeptides therefore include natural biological variants (for example, allelic variants or geographical variations within the species from which the polypeptides are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the LBDG5, LBDG10, LBDG11, LBDG12 and/or LBDG13 polypeptides. Such mutants may include polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code. Typical such substitutions are among

Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic residues Lys and Arg; or among the aromatic residues Phe and Tyr. Particularly preferred are variants in which several, *i.e.* between 5 and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are substituted, deleted or added in any combination. Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions. Such mutants also include polypeptides in which one or more of the amino acid residues includes a substituent group.

Typically, greater than 30% identity between two polypeptides (preferably, over a specified region) is considered to be an indication of functional equivalence. Preferably, functionally equivalent polypeptides of the first aspect of the invention have a degree of sequence identity with the LBDG5, LBDG10, LBDG11, LBDG12 and/or LBDG13 polypeptides, or with active fragments thereof, of greater than 80%. More preferred polypeptides have degrees of identity of greater than 85%, 90%, 95%, 98% or 99%, respectively with the LBDG5, LBDG10, LBDG11, LBDG12 and/or LBDG13 polypeptides, or with active fragments thereof.

Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

The functionally-equivalent polypeptides of the first aspect of the invention may also be polypeptides which have been identified using one or more techniques of structural alignment. For example, the Inpharmatica Genome Threader™ technology that forms one aspect of the search tools used to generate the Biopendium search database may be used (see co-pending International patent application WO 01/67507) to identify polypeptides of presently-unknown function which, while having low sequence identity as compared to the LBDG5, LBDG10, LBDG11, LBDG12 and/or LBDG13 polypeptides, are predicted to have Nuclear Hormone Receptor activity, by virtue of sharing significant structural homology with these sequences.

By "significant structural homology" is meant that the Inpharmatica Genome Threader™ predicts two proteins, or protein regions, to share structural homology with a certainty of

at least 10% more preferably, at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and above. The certainty value of the Inpharmatica Genome Threader™ is calculated as follows. A set of comparisons was initially performed using the Inpharmatica Genome Threader™ exclusively using sequences of known structure. Some of the comparisons were between proteins that were known to be related (on the basis of structure). A neural network was then trained on the basis that it needed to best distinguish between the known relationships and known not-relationships taken from the CATH structure classification (www.biochem.ucl.ac.uk/bsm/cath). This resulted in a neural network score between 0 and 1. However, again as the number of proteins that are related and the number that are unrelated were known, it was possible to partition the neural network results into packets and calculate empirically the percentage of the results that were correct. In this manner, any genuine prediction in the Biopendium search database has an attached neural network score and the percentage confidence is a reflection of how successful the Inpharmatica Genome Threader™ was in the training/testing set.

Structural homologues to the Ligand Binding Domain of LBDG5 should share structural homology with the LBDG5 region. Such structural homologues are predicted to have Nuclear Hormone Receptor Ligand Binding Domain activity by virtue of sharing significant structural homology with this polypeptide sequence.

Structural homologues of LBDG11 should share structural homology with the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain. Such structural homologues are predicted to have Nuclear Hormone Receptor Ligand Binding Domain activity by virtue of sharing significant structural homology with this polypeptide sequence.

The polypeptides of the first aspect of the invention also include fragments of the LBDG5 polypeptide, functional equivalents of the fragments of the LBDG5 polypeptide, and fragments of the functional equivalents of the LBDG5 polypeptides (such as fragments of the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, and the LBDG13 polypeptide), provided that those functional equivalents and fragments retain Nuclear Hormone Receptor activity or have an antigenic determinant in common with the LBDG5 polypeptide.

As used herein, the term "fragment" refers to a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of the LBDG5,

LBDG10, LBDG11, LBDG12 and/or LBDG13 polypeptides, or one of its functional equivalents. The fragments should comprise at least n consecutive amino acids from the sequence and, depending on the particular sequence, n preferably is 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more). Small fragments may form an antigenic determinant.

Preferred polypeptide fragments according to this aspect of the invention are fragments that include a region defined herein as the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG10 polypeptide, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG11 polypeptide the LBDG12 region that has been annotated as a Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG12 polypeptide, and the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG13 polypeptide respectively. These regions are the regions that have been annotated as Nuclear Hormone Receptor Ligand Binding Domains.

For the LBDG5 polypeptide, this region is considered to extend between residue 394 and residue 604. For the LBDG10 polypeptide, this region is considered to extend between residue 413 and residue 603. For the LBDG11 polypeptide, this region is considered to extend between residue 118 and residue 319. For the LBDG12 polypeptide, this region is considered to extend between residue 118 and residue 318. For the LBDG13 polypeptide, this region is considered to extend between residue 117 and residue 317.

Variants of these fragments are included as embodiments of this aspect of the invention, provided that these variants possess activity as a Nuclear Hormone Receptor Ligand Binding Domain.

In one respect, the term "variant" is meant to include extended or truncated versions of this polypeptide fragment.

For extended variants, it is considered highly likely that the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide will fold correctly and show Nuclear Hormone Receptor Ligand Binding Domain activity if

additional residues C terminal and/or N terminal of these boundaries in the polypeptide sequence are included in the polypeptide fragment. For example, an additional 5, 10, 20, 30, 40 or even 50 or more amino acid residues from the LBDG5 polypeptide sequence, the LBDG11 polypeptide sequence, the LBDG12 polypeptide sequence, or the the LBDG13 polypeptide sequence respectively may be included in the polypeptide fragment, without prejudicing the ability of the polypeptide fragment to fold correctly and exhibit Nuclear Hormone Receptor Ligand Binding Domain activity. For truncated variants of the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide, one or more amino acid residues, even 5, 10, 20, 30, 40, 50 or more amino acid residues may be deleted at either or both the C terminus or the N terminus of the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide, respectively. For example, a preferred fragment according to the invention is that encompassing amino acids 100-329 of the LBDG11 polypeptide.

In a second respect, the term "variant" includes homologues of the polypeptide fragments described above, that possess significant sequence homology with the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5, LBDG10, LBDG11, LBDG12 and/or LBDG13 polypeptides or one of their functional equivalents, provided that said variants retain activity as an Nuclear Hormone Receptor Ligand Binding Domain.

Homologues include those polypeptide molecules that possess greater than 80% identity with the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, or the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region. Percentage identity is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

Preferably, variant homologues of polypeptide fragments of this aspect of the invention have a degree of sequence identity with the LBDG5 Nuclear Hormone Receptor Ligand

Binding Domain region, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, or the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region, of greater than 85%. More preferred variant polypeptides have degrees of identity of greater than 90%, 95%, 98% or 99%, respectively with these polypeptide regions, provided that said variants retain activity as a Nuclear Hormone Receptor Ligand Binding Domain. Variant polypeptides also include homologues of the truncated forms of the polypeptide fragments discussed above, provided that said variants retain activity as a Nuclear Hormone Receptor Ligand Binding Domain.

The polypeptide fragments of the first aspect of the invention may be polypeptide fragments that exhibit significant structural homology with the structure of the polypeptide fragment defined by the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, or the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region, for example, as identified by the Inpharmatica Genome Threader™. Accordingly, polypeptide fragments that are structural homologues of the polypeptide fragments defined by LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, or the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region should adopt the same fold as that adopted by this polypeptide fragment, as this fold is defined above.

Such fragments may be "free-standing", *i.e.* not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the fragment of the invention most preferably forms a single continuous region. For instance, certain preferred embodiments relate to a fragment having a pre- and/or pro- polypeptide region fused to the amino terminus of the fragment and/or an additional region fused to the carboxyl

terminus of the fragment. However, several fragments may be comprised within a single larger polypeptide.

Homo and heterodimers of the polypeptide fragments and variants as these are described above are also included in this aspect of the invention. Preferred polypeptide partners in heterodimers of the LBDG5, LBDG10, LBDG11, LBDG12 and LBDG13 polypeptides include polypeptides that contain Nuclear Hormone Receptor Ligand Binding Domains, such as ERa and ERß.

The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies may be employed to isolate or to identify clones expressing the polypeptides of the invention or to purify the polypeptides by affinity chromatography. The antibodies may also be employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')2 and Fv, which are capable of binding to the antigenic determinant in question. Such antibodies thus bind to the polypeptides of the first aspect of the invention.

By "substantially greater affinity" we mean that there is a measurable increase in the affinity for a polypeptide of the invention as compared with the affinity for known Nuclear Hormone Receptors.

Preferably, the affinity is at least 1.5-fold, 2-fold, 5-fold 10-fold, 100-fold, 10^3-fold, 10^4-fold, 10^5-fold, 10^6-fold or greater for a polypeptide of the invention than for known Nuclear Hormone Receptors.

If polyclonal antibodies are desired, a selected mammal, such as a mouse, rabbit, goat or horse, may be immunised with a polypeptide of the first aspect of the invention. The polypeptide used to immunise the animal can be derived by recombinant DNA technology or can be synthesized chemically. If desired, the polypeptide can be

conjugated to a carrier protein. Commonly used carriers to which the polypeptides may be chemically coupled include bovine serum albumin, thyroglobulin and keyhole limpet haemocyanin. The coupled polypeptide is then used to immunise the animal. Serum from the immunised animal is collected and treated according to known procedures, for example by immunoaffinity chromatography.

Monoclonal antibodies to the polypeptides of the first aspect of the invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies using hybridoma technology is well known (see, for example, Kohler, G. and Milstein, C., Nature 256: 495-497 (1975); Kozbor *et al.*, Immunology Today 4: 72 (1983); Cole *et al.*, 77-96 in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985).

Panels of monoclonal antibodies produced against the polypeptides of the first aspect of the invention can be screened for various properties, *i.e.*, for isotype, epitope, affinity, etc. Monoclonal antibodies are particularly useful in purification of the individual polypeptides against which they are directed. Alternatively, genes encoding the monoclonal antibodies of interest may be isolated from hybridomas, for instance by PCR techniques known in the art, and cloned and expressed in appropriate vectors.

Chimeric antibodies, in which non-human variable regions are joined or fused to human constant regions (see, for example, Liu *et al.*, Proc. Natl. Acad. Sci. USA, 84, 3439 (1987)), may also be of use.

The antibody may be modified to make it less immunogenic in an individual, for example by humanisation (see Jones *et al.*, Nature, 321, 522 (1986); Verhoeyen *et al.*, Science, 239: 1534 (1988); Kabat *et al.*, J. Immunol., 147: 1709 (1991); Queen *et al.*, Proc. Natl Acad. Sci. USA, 86, 10029 (1989); Gorman *et al.*, Proc. Natl Acad. Sci. USA, 88: 34181 (1991); and Hodgson *et al.*, Bio/Technology 9: 421 (1991)). The term "humanised antibody", as used herein, refers to antibody molecules in which the CDR amino acids and selected other amino acids in the variable domains of the heavy and/or light chains of a non-human donor antibody have been substituted in place of the equivalent amino acids in a human antibody. The humanised antibody thus closely resembles a human antibody but has the binding ability of the donor antibody.

In a further alternative, the antibody may be a "bispecific" antibody, that is an antibody having two different antigen-binding domains, each domain being directed against a different epitope.

Phage display technology may be utilised to select genes which encode antibodies with binding activities towards the polypeptides of the invention either from repertoires of PCR amplified V-genes of lymphocytes from humans screened for possessing the relevant antibodies, or from naive libraries (McCafferty, J. *et al.*, (1990), Nature 348, 552-554; Marks, J. *et al.*, (1992) Biotechnology 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) Nature 352, 624-628).

Antibodies generated by the above techniques, whether polyclonal or monoclonal, have additional utility in that they may be employed as reagents in immunoassays, radioimmunoassays (RIA) or enzyme-linked immunosorbent assays (ELISA). In these applications, the antibodies can be labelled with an analytically-detectable reagent such as a radioisotope, a fluorescent molecule or an enzyme.

Preferred nucleic acid molecules of the second and third aspects of the invention are those which encode the polypeptide sequences recited in SEQ ID NO:2 (particularly the coding region, nucleotides 61-1905), SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, or SEQ ID NO:10, and functionally equivalent polypeptides, including active fragments of the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, and the LBDG13 polypeptide such as a fragment including the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide sequence, a fragment including the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG10 polypeptide sequence, a fragment including the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG11 polypeptide sequence, a fragment including the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG12 polypeptide sequence, a fragment including the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG13 polypeptide sequence or a homologue thereof and homo- and heterodimers of the LBDG5 polypeptide and the LBDG10 polypeptide and/or its active fragments.

Nucleic acid molecules encompassing these stretches of sequence form a preferred embodiment of this aspect of the invention.

These nucleic acid molecules may be used in the methods and applications described herein. The nucleic acid molecules of the invention preferably comprise at least n consecutive nucleotides from the sequences disclosed herein where, depending on the particular sequence, n is 10 or more (for example, 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

The nucleic acid molecules of the invention also include sequences that are complementary to nucleic acid molecules described above (for example, for antisense or probing purposes).

Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance cDNA, synthetic DNA or genomic DNA. Such nucleic acid molecules may be obtained by cloning, by chemical synthetic techniques or by a combination thereof. The nucleic acid molecules can be prepared, for example, by chemical synthesis using techniques such as solid phase phosphoramidite chemical synthesis, from genomic or cDNA libraries or by separation from an organism. RNA molecules may generally be generated by the *in vitro* or *in vivo* transcription of DNA sequences.

The nucleic acid molecules may be double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The term "nucleic acid molecule" also includes analogues of DNA and RNA, such as those containing modified backbones, and peptide nucleic acids (PNA). The term "PNA", as used herein, refers to an antisense molecule or an anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues, which preferably ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in a cell, where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. *et al.* (1993) Anticancer Drug Des. 8:53-63).

A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:2, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:1. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:4, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:3. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:6, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:5. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:8, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:7. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:10, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:9. These molecules also may have a different sequence which, as a result of the degeneracy of the genetic code, encodes the polypeptide SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or an active fragment of the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide, such as a fragment including the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 394 and residue 604 of the LBDG5 polypeptide sequence. In SEQ ID NO:1 the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 1240 to nucleotide 1872. Nucleic acid molecules encompassing this stretch of sequence, and variants and homologues of this sequence, form a preferred embodiment of this aspect of the invention. An example of such a homologue is the fragment of LBDG11 that encompasses amino acids 100-329 of the full length LBDG11 polypeptide.

The LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 413 and residue 603 of the LBDG10 polypeptide sequence. In

SEQ ID NO:3 the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 1237 to nucleotide 1809. The LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 118 and residue 319 of the LBDG11 polypeptide sequence. In SEQ ID NO:5 the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 352 to nucleotide 957. The LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 118 and residue 318 of the LBDG12 polypeptide sequence. In SEQ ID NO:7 the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 352 to nucleotide 954. The LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 117 and residue 317 of the LBDG13 polypeptide sequence. In SEQ ID NO:9 the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 349 to nucleotide 951. Nucleic acid molecules encompassing these stretches of sequence, and homologues of these sequences, form a preferred embodiment of this aspect of the invention.

In one respect, the term "variant" is meant to include extended or truncated versions of this polynucleotide fragment.

For extended variants, it is considered highly likely that the polynucleotide sequences which encode the Nuclear Hormone Receptor Ligand Binding Domain regions of the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide will fold correctly and show Nuclear Hormone Receptor Ligand Binding Domain activity if additional residues 5' and/or 3' nucleotides terminal of these boundaries in the polynucleotide sequence are included in the polynucleotide fragment. For example, an additional 5, 15, 60, 90, 120 or even 150 or more nucleotides from the LBDG5 sequence, the LBDG11 sequence, the LBDG12 sequence, or the LBDG13 sequence respectively.

Such nucleic acid molecules that encode the polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10 may include, but are not limited to, the coding sequence for the mature polypeptide by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or

secretory sequence, such as a pro-, pre- or prepro- polypeptide sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with further additional, non-coding sequences, including non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription (including termination signals), ribosome binding and mRNA stability. The nucleic acid molecules may also include additional sequences which encode additional amino acids, such as those which provide additional functionalities.

The nucleic acid molecules of the second and third aspects of the invention may also encode the fragments or the functional equivalents of the polypeptides and fragments of the first aspect of the invention.

As discussed above, a preferred fragment of the LBDG5 polypeptide is a fragment including the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The Nuclear Hormone Receptor Ligand Binding Domain region is encoded by a nucleic acid molecule including nucleotide 1240 to nucleotide 1872 of SEQ ID NO:1. A preferred fragment of the LBDG10 polypeptide is a fragment including the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region is encoded by a nucleic acid molecule including nucleotide 1237 to nucleotide 1809 of SEQ ID NO:3. A preferred fragment of the LBDG11 polypeptide is a fragment including the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region is encoded by a nucleic acid molecule including nucleotide 352 to nucleotide 957 of SEQ ID NO:5. A preferred fragment of the LBDG12 polypeptide is a fragment including the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region is encoded by a nucleic acid molecule including nucleotide 352 to nucleotide 954 of SEQ ID NO:7. A preferred fragment of the LBDG13 polypeptide is a fragment including the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region is encoded by a nucleic acid molecule including nucleotide 349 to nucleotide 951 of SEQ ID NO:9.

Functionally-equivalent nucleic acid molecules according to the invention may be naturally-occurring variants such as a naturally-occurring allelic variant, or the molecules may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the nucleic acid molecule may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells or organisms.

In one respect, the term "variant" is meant to include extended or truncated versions of this polynucleotide fragment. For extended variants, it is considered highly likely that the polynucleotide sequences which encode the Nuclear Hormone Receptor Ligand Binding Domain regions of the polypeptides of the invention will fold correctly and show Nuclear Hormone Receptor Ligand Binding Domain activity if additional residues 5' and/or 3' nucleotides terminal of these boundaries in the polynucleotide sequence are included in the polynucleotide fragment. For example, an additional 5, 15, 60, 90, 120 or even 150 or more nucleotides from the polypeptide sequences may be included.

Among variants in this regard are variants that differ from the aforementioned nucleic acid molecules by nucleotide substitutions, deletions or insertions. The substitutions, deletions or insertions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or insertions.

The nucleic acid molecules of the invention can also be engineered, using methods generally known in the art, for a variety of reasons, including modifying the cloning, processing, and/or expression of the gene product (the polypeptide). DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides are included as techniques which may be used to engineer the nucleotide sequences. Site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations and so forth.

Nucleic acid molecules which encode a polypeptide of the first aspect of the invention may be ligated to a heterologous sequence so that the combined nucleic acid molecule encodes a fusion protein. Such combined nucleic acid molecules are included within the second or third aspects of the invention. For example, to screen peptide libraries for inhibitors of the activity of the polypeptide, it may be useful to express, using such a

combined nucleic acid molecule, a fusion protein that can be recognised by a commercially-available antibody. A fusion protein may also be engineered to contain a cleavage site located between the sequence of the polypeptide of the invention and the sequence of a heterologous protein so that the polypeptide may be cleaved and purified away from the heterologous protein.

The nucleic acid molecules of the invention also include antisense molecules that are partially complementary to nucleic acid molecules encoding polypeptides of the present invention and that therefore hybridize to the encoding nucleic acid molecules (hybridization). Such antisense molecules, such as oligonucleotides, can be designed to recognise, specifically bind to and prevent transcription of a target nucleic acid encoding a polypeptide of the invention, as will be known by those of ordinary skill in the art (see, for example, Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee *et al.*, Nucleic Acids Res 6, 3073 (1979); Cooney *et al.*, Science 241, 456 (1988); Dervan *et al.*, Science 251, 1360 (1991).

The term "hybridization" as used here refers to the association of two nucleic acid molecules with one another by hydrogen bonding. Typically, one molecule will be fixed to a solid support and the other will be free in solution. Then, the two molecules may be placed in contact with one another under conditions that favour hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase molecule to the solid support (Denhardt's reagent or BLOTTO); the concentration of the molecules; use of compounds to increase the rate of association of molecules (dextran sulphate or polyethylene glycol); and the stringency of the washing conditions following hybridization (see Sambrook *et al.* [*supra*]).

The inhibition of hybridization of a completely complementary molecule to a target molecule may be examined using a hybridization assay, as known in the art (see, for example, Sambrook *et al.* [*supra*]). A substantially homologous molecule will then compete for and inhibit the binding of a completely homologous molecule to the target molecule under various conditions of stringency, as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol.

152:507-511).

"Stringency" refers to conditions in a hybridization reaction that favour the association of very similar molecules over association of molecules that differ. High stringency hybridisation conditions are defined as overnight incubation at 42°C in a solution comprising 50% formamide, 5XSSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardts solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1X SSC at approximately 65°C. Low stringency conditions involve the hybridisation reaction being carried out at 35°C (see Sambrook *et al.* [*supra*]). Preferably, the conditions used for hybridization are those of high stringency.

Preferred embodiments of this aspect of the invention are nucleic acid molecules that are at least 70% identical over their entire length to a nucleic acid molecule encoding the LBDG5 polypeptide (SEQ ID NO:2), the LBDG10 polypeptide (SEQ ID NO:4), the LBDG11 polypeptide (SEQ ID NO:6), the LBDG12 polypeptide (SEQ ID NO:8), the LBDG13 polypeptide (SEQ ID NO:10), and nucleic acid molecules that are substantially complementary to such nucleic acid molecules. A preferred active fragment is a fragment that includes an LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide sequence. A further preferred active fragment is a fragment that includes an LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG10 polypeptide sequence. A further preferred active fragment is a fragment that includes an LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG11 polypeptide sequence. A further preferred active fragment is a fragment that includes an LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG12 polypeptide sequence. A further preferred active fragment is a fragment that includes an LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG13 polypeptide sequence. Accordingly, preferred nucleic acid molecules include those that are at least 70% identical over their entire length to a nucleic acid molecule encoding the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide sequence, the LBDG10 polypeptide sequence, the LBDG11 polypeptide sequence, the LBDG12 polypeptide sequence, and the LBDG13 polypeptide sequence.

Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/).

Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to the nucleic acid molecule having the sequence given in SEQ ID NO:1, to a region including nucleotides 1240 to 1872 of SEQ ID NO:1, the nucleic acid molecule having the sequence given in SEQ ID NO:3, to a region including nucleotides 1237 to 1809 of SEQ ID NO:3, the nucleic acid molecule having the sequence given in SEQ ID NO:5, to a region including nucleotides 352 to 957 of SEQ ID NO:5, the nucleic acid molecule having the sequence given in SEQ ID NO:7, to a region including nucleotides 352 to 954 of SEQ ID NO:7, the nucleic acid molecule having the sequence given in SEQ ID NO:9, to a region including nucleotides 349-951 of SEQ ID NO:9, or a nucleic acid molecule that is complementary to any one of these regions of nucleic acid. In this regard, nucleic acid molecules at least 90%, preferably at least 95%, more preferably at least 98% or 99% identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are nucleic acid molecules that encode polypeptides which retain substantially the same biological function or activity as the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide.

The invention also provides a process for detecting a nucleic acid molecule of the invention, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting any such duplexes that are formed.

As discussed additionally below in connection with assays that may be utilised according to the invention, a nucleic acid molecule as described above may be used as a hybridization probe for RNA, cDNA or genomic DNA, in order to isolate full-length cDNAs and genomic clones encoding the LBDG5 polypeptide or its functional equivalents and to isolate cDNA and genomic clones of homologous or orthologous genes that have a high sequence similarity to the gene encoding this polypeptide.

In this regard, the following techniques, among others known in the art, may be utilised and are discussed below for purposes of illustration. Methods for DNA sequencing and

analysis are well known and are generally available in the art and may, indeed, be used to practise many of the embodiments of the invention discussed herein. Such methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proof-reading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD). Preferably, the sequencing process may be automated using machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), the Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

One method for isolating a nucleic acid molecule encoding a polypeptide with an equivalent function to that of the LBDG5 polypeptide, particularly with an equivalent function to the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide, is to probe a genomic or cDNA library with a natural or artificially-designed probe using standard procedures that are recognised in the art (see, for example, "Current Protocols in Molecular Biology", Ausubel *et al.* (eds). Greene Publishing Association and John Wiley Interscience, New York, 1989,1992). Probes comprising at least 15, preferably at least 30, and more preferably at least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate encoding gene (SEQ ID NO:1) are particularly useful probes. In the case of SEQ ID NO:1, particularly suitable probes may be selected from nucleotides 1240 to1872. In the case of SEQ ID NO:3, particularly suitable probes may be selected from nucleotides 1237 to1809. In the case of SEQ ID NO:5, particularly suitable probes may be selected from nucleotides 352 to 957. In the case of SEQ ID NO:7, particularly suitable probes may be selected from nucleotides 352 to 954. In the case of SEQ ID NO:9, particularly suitable probes may be selected from nucleotides 349 to 951.

Such probes may be labelled with an analytically-detectable reagent to facilitate their identification. Useful reagents include, but are not limited to, radioisotopes, fluorescent dyes and enzymes that are capable of catalysing the formation of a detectable product. Using these probes, the ordinarily skilled artisan will be capable of isolating complementary copies of genomic DNA, cDNA or RNA polynucleotides encoding

proteins of interest from human, mammalian or other animal sources and screening such sources for related sequences, for example, for additional members of the family, type and/or subtype.

In many cases, isolated cDNA sequences will be incomplete, in that the region encoding the polypeptide will be cut short, normally at the 5' end. Several methods are available to obtain full length cDNAs, or to extend short cDNAs. Such sequences may be extended utilising a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed is based on the method of Rapid Amplification of cDNA Ends (RACE; see, for example, Frohman *et al.*, Proc. Natl. Acad. Sci. USA (1988) 85: 8998-9002). Recent modifications of this technique, exemplified by the MarathonTM technology (Clontech Laboratories Inc.), for example, have significantly simplified the search for longer cDNAs. A slightly different technique, termed "restriction-site" PCR, uses universal primers to retrieve unknown nucleic acid sequence adjacent a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Inverse PCR may also be used to amplify or to extend sequences using divergent primers based on a known region (Triglia, T., *et al.* (1988) Nucleic Acids Res. 16:8186). Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. *et al.* (1991) PCR Methods Applic. 1: 111-119). Another method which may be used to retrieve unknown sequences is that of Parker, J.D. *et al.* (1991); Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences that contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

In one embodiment of the invention, the nucleic acid molecules of the present invention may be used for chromosome localisation. In this technique, a nucleic acid molecule is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important step in the confirmatory correlation of those sequences with the gene-associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationships between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleic acid molecule may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

The nucleic acid molecules of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the polypeptide in tissues by detection of the mRNAs that encode them. These techniques include *in situ* hybridization techniques and nucleotide amplification techniques, such as PCR. Results from these studies provide an indication of the normal functions of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable insights into the role of mutant polypeptides in disease. Such inappropriate expression may be of a temporal, spatial or quantitative nature.

Gene silencing approaches may also be undertaken to down-regulate endogenous expression of a gene encoding a polypeptide of the invention. RNA interference (RNAi) (Elbashir, SM *et al.*, Nature 2001, 411, 494-498) is one method of sequence specific post-transcriptional gene silencing that may be employed. Short dsRNA oligonucleotides are

synthesised *in vitro* and introduced into a cell. The sequence specific binding of these dsRNA oligonucleotides triggers the degradation of target mRNA, reducing or ablating target protein expression.

Efficacy of the gene silencing approaches assessed above may be assessed through the measurement of polypeptide expression (for example, by Western blotting), and at the RNA level using TaqMan-based methodologies.

The vectors of the present invention comprise nucleic acid molecules of the invention and may be cloning or expression vectors. The host cells of the invention, which may be transformed, transfected or transduced with the vectors of the invention may be prokaryotic or eukaryotic.

The polypeptides of the invention may be prepared in recombinant form by expression of their encoding nucleic acid molecules in vectors contained within a host cell. Such expression methods are well known to those of skill in the art and many are described in detail by Sambrook *et al.* (supra) and Fernandez & Hoeffler (1998, eds. "Gene expression systems. Using nature for the art of expression". Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto).

Generally, any system or vector that is suitable to maintain, propagate or express nucleic acid molecules to produce a polypeptide in the required host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those described in Sambrook *et al.*, (supra). Generally, the encoding gene can be placed under the control of a control element such as a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator, so that the DNA sequence encoding the desired polypeptide is transcribed into RNA in the transformed host cell.

Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic

elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid.

Particularly suitable expression systems include microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (for example, baculovirus); plant cell systems transformed with virus expression vectors (for example, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (for example, Ti or pBR322 plasmids); or animal cell systems. Cell-free translation systems can also be employed to produce the polypeptides of the invention.

Introduction of nucleic acid molecules encoding a polypeptide of the present invention into host cells can be effected by methods described in many standard laboratory manuals, such as Davis *et al.*, Basic Methods in Molecular Biology (1986) and Sambrook *et al.*, [*supra*]. Particularly suitable methods include calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection (see Sambrook *et al.*, 1989 [*supra*]; Ausubel *et al.*, 1991 [*supra*]; Spector, Goldman & Leinwald, 1998). In eukaryotic cells, expression systems may either be transient (for example, episomal) or permanent (chromosomal integration) according to the needs of the system.

The encoding nucleic acid molecule may or may not include a sequence encoding a control sequence, such as a signal peptide or leader sequence, as desired, for example, for secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals. Leader sequences can be removed by the bacterial host in post-translational processing.

In addition to control sequences, it may be desirable to add regulatory sequences that allow for regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory sequences are those which cause the expression of a gene to be increased or decreased in response to a chemical or physical stimulus, including the

presence of a regulatory compound or to various temperature or metabolic conditions. Regulatory sequences are those non-translated regions of the vector, such as enhancers, promoters and 5' and 3' untranslated regions. These interact with host cellular proteins to carry out transcription and translation. Such regulatory sequences may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript phagemid (Stratagene, LaJolla, CA) or pSportl™ plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (for example, heat shock, RUBISCO and storage protein genes) or from plant viruses (for example, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

An expression vector is constructed so that the particular nucleic acid coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the regulatory sequences being such that the coding sequence is transcribed under the "control" of the regulatory sequences, *i.e.*, RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence. In some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; *i.e.*, to maintain the reading frame.

The control sequences and other regulatory sequences may be ligated to the nucleic acid coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector that already contains the control sequences and an appropriate restriction site.

For long-term, high-yield production of a recombinant polypeptide, stable expression is preferred. For example, cell lines which stably express the polypeptide of interest may be transformed using expression vectors which may contain viral origins of replication

and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalised cell lines available from the American Type Culture Collection (ATCC) including, but not limited to, Chinese hamster ovary (CHO), HeLa, baby hamster kidney (BHK), monkey kidney (COS), C127, 3T3, BHK, HEK 293, Bowes melanoma and human hepatocellular carcinoma (for example Hep G2) cells and a number of other cell lines.

In the baculovirus system, the materials for baculovirus/insect cell expression systems are commercially available in kit form from, inter alia, Invitrogen, San Diego CA (the "MaxBac" kit). These techniques are generally known to those skilled in the art and are described fully in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987). Particularly suitable host cells for use in this system include insect cells such as Drosophila S2 and Spodoptera Sf9 cells.

There are many plant cell culture and whole plant genetic expression systems known in the art. Examples of suitable plant cellular genetic expression systems include those described in US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, (1991) Phytochemistry 30, 3861-3863.

In particular, all plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be utilised, so that whole plants are recovered which contain the transferred gene. Practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and other trees, legumes and vegetables.

Examples of particularly preferred bacterial host cells include *streptococci*, *staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells.

Examples of particularly suitable host cells for fungal expression include yeast cells (for example, *S. cerevisiae*) and *Aspergillus* cells.

Any number of selection systems are known in the art that may be used to recover transformed cell lines. Examples include the herpes simplex virus thymidine kinase (Wigler, M. *et al.* (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. *et al.* (1980) Cell 22:817-23) genes that can be employed in tk^- or $aprt^\pm$ cells, respectively.

Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dihydrofolate reductase (DHFR) that confers resistance to methotrexate (Wigler, M. *et al.* (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. *et al* (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. Additional selectable genes have been described, examples of which will be clear to those of skill in the art.

Although the presence or absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the relevant sequence is inserted within a marker gene sequence, transformed cells containing the appropriate sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a polypeptide of the invention under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain a nucleic acid sequence encoding a polypeptide of the invention and which express said polypeptide may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassays, for example, fluorescence activated cell sorting (FACS) or immunoassay techniques (such as the enzyme-linked immunosorbent assay [ELISA] and radioimmunoassay [RIA]), that

include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein (see Hampton, R. *et al.* (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. *et al.* (1983) J. Exp. Med, 158, 1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labelled hybridization or PCR probes for detecting sequences related to nucleic acid molecules encoding polypeptides of the present invention include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled polynucleotide. Alternatively, the sequences encoding the polypeptide of the invention may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesise RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH)).

Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes and fluorescent, chemiluminescent or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Nucleic acid molecules according to the present invention may also be used to create transgenic animals, particularly rodent animals. Such transgenic animals form a further aspect of the present invention. This may be done locally by modification of somatic cells, or by germ line therapy to incorporate heritable modifications. Such transgenic animals may be particularly useful in the generation of animal models for drug molecules effective as modulators of the polypeptides of the present invention.

The polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. High performance liquid chromatography is particularly useful for purification. Well known techniques for refolding proteins may be

employed to regenerate an active conformation when the polypeptide is denatured during isolation and or purification.

Specialised vector constructions may also be used to facilitate purification of proteins, as desired, by joining sequences encoding the polypeptides of the invention to a nucleotide sequence encoding a polypeptide domain that will facilitate purification of soluble proteins. Examples of such purification-facilitating domains include metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the polypeptide of the invention may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing the polypeptide of the invention fused to several histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification by IMAC (immobilised metal ion affinity chromatography as described in Porath, J. *et al.* (1992) Prot. Exp. Purif. 3: 263-281) while the thioredoxin or enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. *et al.* (DNA Cell Biol. 199312:441-453).

If the polypeptide is to be expressed for use in screening assays, generally it is preferred that it be produced at the surface of the host cell in which it is expressed. In this event, the host cells may be harvested prior to use in the screening assay, for example using techniques such as fluorescence activated cell sorting (FACS) or immunoaffinity techniques. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the expressed polypeptide. If polypeptide is produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

The polypeptide of the invention can be used to screen libraries of compounds in any of a variety of drug screening techniques. Such compounds may activate (agonise) or inhibit (antagonise) the level of expression of the gene or the activity of the polypeptide of the invention and form a further aspect of the present invention. Preferred compounds are

effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

Agonist or antagonist compounds may be isolated from, for example, cells, cell-free preparations, chemical libraries or natural product mixtures. These agonists or antagonists may be natural or modified substrates, ligands, enzymes, receptors or structural or functional mimetics. For a suitable review of such screening techniques, see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991).

Compounds that are most likely to be good agonists and antagonists are molecules that bind to the polypeptide of the invention modulating the biological effects of the polypeptide upon binding to it. Potential agonists and antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to the polypeptide of the invention and thereby activate, inhibit or extinguish its activity. In this fashion, binding of the polypeptide to normal cellular binding molecules may be modulated, such that the normal biological activity of the polypeptide is enhanced, or decreased for therapeutic utility.

The functional effects for the nuclear receptor gene family are typically modified by ligands, either with transcription being upregulated or downregulated by addition of a ligand. These ligands may either be therapeutic agents themselves, or form the basis for novel therapeutic agents. The set of ligands that known nuclear receptor ligand binding domains interact with, include, but are not limited by, steroid, fatty acid, vitamins, and similar molecules. Modulation of the transcriptional activity of nuclear receptors by such ligand is thus indicative of their functional activity.

Given the known natural ligands for established nuclear receptors, one can search for chemically similar compounds from the Available Chemical Directory (ACD), using standard compound similarity searching tools (supplied by MDL, San Leandro, CA). We additionally generated a number of 3D pharmacophores for known nuclear receptors, using our Chematica suite of software (see PCT/GB01/0277), and used these to search for compounds displaying potentially nuclear receptor specific interaction properties. By following this method, which can be achieved alternatively using a variety of different 2D and 3D searching approaches, we had generated a set of 2,469 ligands, as probes for

nuclear receptor function.

The polypeptide of the invention that is employed in such a screening technique may be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. In general, such screening procedures may involve using appropriate cells or cell membranes that express the polypeptide that are contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. The functional response of the cells contacted with the test compound is then compared with control cells that were not contacted with the test compound. Such an assay may assess whether the test compound results in a signal generated by activation of the polypeptide, using an appropriate detection system. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of the test compound is observed.

A preferred method for identifying an agonist or antagonist compound of a polypeptide of the present invention comprises determining whether the compound binds to and activates or inhibits the polypeptide by measuring the level of a signal generated from the interaction of the compound with the polypeptide.

Methods for generating detectable signals in the types of assays described herein will be known to those of skill in the art. A particular example is cotransfecting a construct expressing a polypeptide according to the invention, or a fragment such as the LBD, in fusion with the GAL4 DNA binding domain, into a cell together with a reporter plasmid, an example of which is pFR-Luc (Stratagene Europe, Amsterdam, The Netherlands). This particular plasmid contains a synthetic promoter with five tandem repeats of GAL4 binding sites that control the expression of the luciferase gene. When a potential ligand is added to the cells, it will bind the GAL4-polypeptide fusion and induce transcription of the luciferase gene. The level of the luciferase expression can be monitored by its activity using a luminescence reader (see, for example, Lehman *et al.* JBC 270, 12953, 1995; Pawar *et al.* JBC, 277, 39243, 2002).

A further preferred method for identifying an agonist or antagonist of a polypeptide of the invention comprises:

(a) contacting a labelled or unlabeled compound with the polypeptide immobilized on any solid support (for example beads, plates, matrix support, chip) and detection of

the compound by measuring the label or the presence of the compound itself; or

(b) contacting a cell expressing on the surface thereof the polypeptide, by means of artificially anchoring it to the cell membrane, or by constructing a chimeric receptor being associated with a second component capable of providing a detectable signal in response to the binding of a compound to the polypeptide, with a compound to be screened under conditions to permit binding to the polypeptide; and

(c) determining whether the compound binds to and activates or inhibits the polypeptide by comparing the level of a signal generated from the interaction of the compound with the polypeptide with the level of a signal in the absence of the compound.

For example, a method such as FRET detection of a ligand bound to the polypeptide in the presence of peptide co-activators (Norris *et al.*, Science 285, 744, 1999) might be used.

In further preferred embodiments, the general methods that are described above may further comprise conducting the identification of agonist or antagonist in the presence of labelled or unlabelled ligand for the polypeptide.

In another embodiment of the method for identifying agonist or antagonist of a polypeptide of the present invention comprises:

determining the inhibition of binding of a ligand to the polypeptide of the invention on any solid or cellular surface thereof, in the presence of a candidate compound under conditions to permit binding to the polypeptide, and determining the amount of ligand bound to the polypeptide. A compound capable of causing reduction of binding of a ligand is considered to be a competitor which may act as an agonist or antagonist. Preferably the ligand is labelled.

More particularly, a method of screening for a polypeptide antagonist or agonist compound comprises the steps of:

(a) incubating a labelled ligand with a polypeptide according to the invention on any solid support (for example beads, plates, matrix support, chip) or the cell surface, or a cell membrane containing a polypeptide of the invention.

(b) measuring the amount of labelled ligand bound to the polypeptide on the solid

support, whole cell or the cell membrane;

(c) adding a candidate compound to a mixture of labelled ligand and immobilized polypeptide on the solid support, the whole cell or the cell membrane of step (a) and allowing the mixture to attain equilibrium;

(d) measuring the amount of labelled ligand bound to the immobilized polypeptide or the whole cell or the cell membrane after step (c); and

(e) comparing the difference in the labelled ligand bound in step (b) and (d), such that the compound which causes the reduction in binding in step (d) is considered to be an agonist or antagonist.

The polypeptides may be found to modulate a variety of physiological and pathological processes in a dose-dependent manner in the above-described assays. Thus, the "functional equivalents" of the polypeptides of the invention include polypeptides that exhibit any of the same modulatory activities in the above-described assays in a dose-dependent manner. Although the degree of dose-dependent activity need not be identical to that of the polypeptides of the invention, preferably the "functional equivalents" will exhibit substantially similar dose-dependence in a given activity assay compared to the polypeptides of the invention.

In certain of the embodiments described above, simple binding assays may be used, in which the adherence of a test compound to a surface bearing the polypeptide is detected by means of a label directly or indirectly associated with the test compound or in an assay involving competition with a labelled competitor. In another embodiment, competitive drug screening assays may be used, in which neutralising antibodies that are capable of binding the polypeptide specifically compete with a test compound for binding. In this manner, the antibodies can be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide.

Assays may also be designed to detect the effect of added test compounds on the production of mRNA encoding the polypeptide in cells. For example, an ELISA may be constructed that measures secreted or cell-associated levels of polypeptide using monoclonal or polyclonal antibodies by standard methods known in the art, and this can be used to search for compounds that may inhibit or enhance the production of the

polypeptide from suitably manipulated cells or tissues. The formation of binding complexes between the polypeptide and the compound being tested may then be measured.

Assay methods that are also included within the terms of the present invention are those that involve the use of the genes and polypeptides of the invention in overexpression or ablation assays. Such assays involve the manipulation of levels of these genes/polypeptides in cells and assessment of the impact of this manipulation event on the physiology of the manipulated cells. For example, such experiments reveal details of signalling and metabolic pathways in which the particular genes/polypeptides are implicated, generate information regarding the identities of polypeptides with which the studied polypeptides interact and provide clues as to methods by which related genes and proteins are regulated.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest (see International patent application WO84/03564). In this method, large numbers of different small test compounds are synthesised on a solid substrate, which may then be reacted with the polypeptide of the invention and washed. One way of immobilising the polypeptide is to use non-neutralising antibodies. Bound polypeptide may then be detected using methods that are well known in the art. Purified polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques.

Examples of suitable assays for the identification of agonists or antagonists of the polypeptides of the invention are described in Rosen *et al.,* Curr. Opin. Drug Discov. Devel. 2003 6(2):224-30.

The polypeptide of the invention may be used to identify membrane-bound or soluble receptors, through standard receptor binding techniques that are known in the art, such as ligand binding and crosslinking assays in which the polypeptide is labelled with a radioactive isotope, is chemically modified, or is fused to a peptide sequence that facilitates its detection or purification, and incubated with a source of the putative receptor (for example, a composition of cells, cell membranes, cell supernatants, tissue extracts, or bodily fluids). The efficacy of binding may be measured using biophysical techniques such as surface plasmon resonance (supplied by Biacore AB, Uppsala,

Sweden) and spectroscopy. Using such assays, it will be possible to investigate the degree to which polypeptides of the invention dimerise or oligomerise with other polypeptide molecules.

Binding assays may be used for the purification and cloning of the receptor, but may also identify agonists and antagonists of the polypeptide, that compete with the binding of the polypeptide to its receptor. Standard methods for conducting screening assays are well understood in the art.

The invention also includes a screening kit useful in the methods for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, that are described above.

The invention includes the agonists, antagonists, ligands, receptors, substrates and enzymes, and other compounds which modulate the activity or antigenicity of the polypeptide of the invention discovered by the methods that are described above.

The invention also provides pharmaceutical compositions comprising a polypeptide, nucleic acid, ligand or compound of the invention in combination with a suitable pharmaceutical carrier. These compositions may be suitable as therapeutic or diagnostic reagents, as vaccines, or as other immunogenic compositions, as outlined in detail below.

According to the terminology used herein, a composition containing a polypeptide, nucleic acid, ligand or compound [X] is "substantially free of" impurities [herein, Y] when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95%, 98% or even 99% by weight.

The pharmaceutical compositions should preferably comprise a therapeutically effective amount of the polypeptide, nucleic acid molecule, ligand, or compound of the invention. The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent needed to treat, ameliorate, or prevent a targeted disease or condition, or to exhibit a detectable therapeutic or preventative effect. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, for example, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and

routes for administration in humans.

The precise effective amount for a human subject will depend upon the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. This amount can be determined by routine experimentation and is within the judgement of the clinician. Generally, an effective dose will be from 0.01 mg/kg to 50 mg/kg, preferably 0.05 mg/kg to 10 mg/kg. Compositions may be administered individually to a patient or may be administered in combination with other agents, drugs or hormones.

A pharmaceutical composition may also contain a pharmaceutically acceptable carrier, for administration of a therapeutic agent. Such carriers include antibodies and other polypeptides, genes and other therapeutic agents such as liposomes, provided that the carrier does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers and inactive virus particles.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulphates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may additionally contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such compositions. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be

treated.

The pharmaceutical compositions utilised in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal or transcutaneous applications (for example, see WO98/20734), subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, intravaginal or rectal means. Gene guns or hyposprays may also be used to administer the pharmaceutical compositions of the invention. Typically, the therapeutic compositions may be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared.

Direct delivery of the compositions will generally be accomplished by injection, subcutaneously, intraperitoneally, intravenously or intramuscularly, or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Dosage treatment may be a single dose schedule or a multiple dose schedule.

If the activity of the polypeptide of the invention is in excess in a particular disease state, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as described above, along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as by blocking the binding of ligands, substrates, enzymes, receptors, or by inhibiting a second signal, and thereby alleviating the abnormal condition. Preferably, such antagonists are antibodies. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as described previously.

In another approach, soluble forms of the polypeptide that retain binding affinity for the ligand, substrate, enzyme, receptor, in question, may be administered. Typically, the polypeptide may be administered in the form of fragments that retain the relevant portions.

In an alternative approach, expression of the gene encoding the polypeptide can be inhibited using expression blocking techniques, such as the use of antisense nucleic acid molecules (as described above), either internally generated or separately administered. Modifications of gene expression can be obtained by designing complementary

sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions (signal sequence, promoters, enhancers and introns) of the gene encoding the polypeptide. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. *et al.* (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Such oligonucleotides may be administered or may be generated in situ from expression *in vivo*.

In addition, expression of the polypeptide of the invention may be prevented by using ribozymes specific to its encoding mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, *et al.*, Curr. Opin. Struct. Biol (1996) 6(4), 527-33). Synthetic ribozymes can be designed to specifically cleave mRNAs at selected positions thereby preventing translation of the mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the ribozymes may be synthesised with non-natural backbones, for example, 2'-O-methyl RNA, to provide protection from ribonuclease degradation and may contain modified bases.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine and butosine, as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine and uridine which are not as easily recognised by endogenous endonucleases.

For treating abnormal conditions related to an under-expression of the polypeptide of the invention and its activity, several approaches are also available. One approach comprises

administering to a subject a therapeutically effective amount of a compound that activates the polypeptide, *i.e.*, an agonist as described above, to alleviate the abnormal condition. Alternatively, a therapeutic amount of the polypeptide in combination with a suitable pharmaceutical carrier may be administered to restore the relevant physiological balance of polypeptide.

Gene therapy may be employed to effect the endogenous production of the polypeptide by the relevant cells in the subject. Gene therapy is used to treat permanently the inappropriate production of the polypeptide by replacing a defective gene with a corrected therapeutic gene.

Gene therapy of the present invention can occur *in vivo* or *ex vivo*. *Ex vivo* gene therapy requires the isolation and purification of patient cells, the introduction of a therapeutic gene and introduction of the genetically altered cells back into the patient. In contrast, in vivo gene therapy does not require isolation and purification of a patient's cells.

The therapeutic gene is typically "packaged" for administration to a patient. Gene delivery vehicles may be non-viral, such as liposomes, or replication-deficient viruses, such as adenovirus as described by Berkner, K.L., in Curr. Top. Microbiol. Immunol., 158, 39-66 (1992) or adeno-associated virus (AAV) vectors as described by Muzyczka, N., in Curr. Top. Microbiol. Immunol., 158, 97-129 (1992) and U.S. Patent No. 5,252,479. For example, a nucleic acid molecule encoding a polypeptide of the invention may be engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding the polypeptide, such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells in vivo and expression of the polypeptide in vivo (see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics (1996), T Strachan and A P Read, BIOS Scientific Publishers Ltd).

Another approach is the administration of "naked DNA" in which the therapeutic gene is directly injected into the bloodstream or muscle tissue.

In situations in which the polypeptides or nucleic acid molecules of the invention are disease-causing agents, the invention provides that they can be used in vaccines to raise antibodies against the disease causing agent.

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection). Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with pharmaceutically-acceptable carriers as described above, which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, and other pathogens.

Since polypeptides may be broken down in the stomach, vaccines comprising polypeptides are preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents.

The vaccine formulations of the invention may be presented in unit-dose or multi-dose containers. For example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Genetic delivery of antibodies that bind to polypeptides according to the invention may also be effected, for example, as described in International patent application WO98/55607.

The technology referred to as jet injection (see, for example, www.powderject.com) may also be useful in the formulation of vaccine compositions.

A number of suitable methods for vaccination and vaccine delivery systems are described in International patent application WO00/29428.

This invention also relates to the use of nucleic acid molecules according to the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the nucleic acid molecules of the invention which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acid molecules for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), or other amplification techniques (see Saiki *et al.*, Nature, 324, 163-166 (1986); Bej, *et al.*, Crit. Rev. Biochem. Molec. Biol., 26, 301-334 (1991); Birkenmeyer *et al.*, J. Virol. Meth., 35, 117-126 (1991); Van Brunt, J., Bio/Technology, 8, 291-294 (1990)) prior to analysis.

In one embodiment, this aspect of the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to the invention and comparing said level of expression to a control level, wherein a level that is different to said control level is indicative of disease. The method may comprise the steps of:

a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule of the invention and the probe;

b) contacting a control sample with said probe under the same conditions used in step a);

c) and detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

Using such methods, it has been found that LBDG11 is up-regulated 2 fold in prostate cancer when compared to the normal sample; a smaller increase in expression was observed in breast cancer compared to control. Expression was down-regulated 2-fold in cervical cancer compared to normal control. Thus, in general, an increase in expression is observed in a variety of cancerous cells. The elevation of LBDG11 transcript in the various cancer samples indicates a role for this nuclear receptor in cancer and points to the development of the identification of therapeutic strategies for the treatment of cancer. Agonists or antagonists against LBDG11 are thus likely to be beneficial in the treatment for cancer. Furthermore, identifying the tumour types that possess elevated LBDG11 transcript levels may be of value in diagnostics for particular cancer subtypes and thereby, define the particular therapeutic approach for individual patients.

A further aspect of the invention comprises a diagnostic method comprising the steps of:

a) obtaining a tissue sample from a patient being tested for disease;

b) isolating a nucleic acid molecule according to the invention from said tissue sample; and

c) diagnosing the patient for disease by detecting the presence of a mutation in the nucleic acid molecule which is associated with disease.

To aid the detection of nucleic acid molecules in the above-described methods, an amplification step, for example using PCR, may be included. Suitable probes are discussed in some detail above.

Deletions and insertions can be detected by a change in the size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labelled RNA of the invention or alternatively, labelled antisense DNA sequences of the invention. Perfectly-matched sequences can be distinguished from mismatched duplexes by RNase digestion or by assessing differences in melting temperatures. The presence or absence of the mutation in the patient may be detected by contacting DNA with a nucleic acid probe that hybridises to the DNA under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or

absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation in the corresponding portion of the DNA strand.

Such diagnostics are particularly useful for prenatal and even neonatal testing.

Point mutations and other sequence differences between the reference gene and "mutant" genes can be identified by other well-known techniques, such as direct DNA sequencing or single-strand conformational polymorphism, (see Orita *et al.*, Genomics, 5, 874-879 (1989)). For example, a sequencing primer may be used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabelled nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. Further, point mutations and other sequence variations, such as polymorphisms, can be detected as described above, for example, through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

DNA sequence differences may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (for example, Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton et al., Proc. Natl. Acad. Sci. USA (1985) 85: 4397-4401).

In addition to conventional gel electrophoresis and DNA sequencing, mutations such as microdeletions, aneuploidies, translocations, inversions, can also be detected by in situ analysis (see, for example, Keller *et al.*, DNA Probes, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), that is, DNA or RNA sequences in cells can be analysed for mutations without need for their isolation and/or immobilisation onto a membrane. Fluorescence *in situ* hybridization (FISH) is presently the most commonly applied method and numerous reviews of FISH have appeared (see, for example, Trachuck *et al.*, Science, 250: 559-562 (1990), and Trask *et al.*, Trends, Genet. 7:149-154 (1991)).

In another embodiment of the invention, an array of oligonucleotide probes comprising a nucleic acid molecule according to the invention can be constructed to conduct efficient screening of genetic variants, mutations and polymorphisms. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee *et al.*, Science (1996) 274: 610-613).

In one embodiment, the array is prepared and used according to the methods described in PCT application WO95/11995 (Chee *et al*); Lockhart, D. J. *et al.* (1996) Nat. Biotech. 14: 1675-1680); and Schena, M. *et al.* (1996) Proc. Natl. Acad. Sci. 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support. In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/25116 (Baldeschweiler et al). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and over one million which lends itself to the efficient use of commercially-available instrumentation.

In addition to the methods discussed above, diseases may be diagnosed by methods comprising determining, from a sample derived from a subject, an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

Assay techniques that can be used to determine levels of a polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and are discussed in some detail above (including radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays). This aspect of the invention provides a diagnostic method which comprises the steps of: (a) contacting a ligand as described above with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

Protocols such as ELISA, RIA, and FACS for measuring polypeptide levels may additionally provide a basis for diagnosing altered or abnormal levels of polypeptide expression. Normal or standard values for polypeptide expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably humans, with antibody to the polypeptide under conditions suitable for complex formation The amount of standard complex formation may be quantified by various methods, such as by photometric means.

Antibodies which specifically bind to a polypeptide of the invention may be used for the diagnosis of conditions or diseases characterised by expression of the polypeptide, or in assays to monitor patients being treated with the polypeptides, nucleic acid molecules, ligands and other compounds of the invention. Antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for the polypeptide include methods that utilise the antibody and a label to detect the polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules known in the art may be used, several of which are described above.

Quantities of polypeptide expressed in subject, control and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. Diagnostic assays may be used to distinguish between absence, presence, and excess expression of polypeptide and to monitor regulation of polypeptide levels during therapeutic intervention. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

A diagnostic kit of the present invention may comprise:

(a) a nucleic acid molecule of the present invention;

(b) a polypeptide of the present invention; or

(c) a ligand of the present invention.

In one aspect of the invention, a diagnostic kit may comprise a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to the invention; a second container containing primers useful for amplifying the nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease. The kit may further comprise a third container holding an agent for digesting unhybridised RNA.

In an alternative aspect of the invention, a diagnostic kit may comprise an array of nucleic acid molecules, at least one of which may be a nucleic acid molecule according to the invention.

To detect polypeptide according to the invention, a diagnostic kit may comprise one or more antibodies that bind to a polypeptide according to the invention; and a reagent useful for the detection of a binding reaction between the antibody and the polypeptide.

Such kits will be of use in diagnosing a disease or susceptibility to diseases in which Nuclear Hormone Receptors are implicated, particularly cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, uterus, prostate, cervical, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposis' sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, hypotension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, Parkinson's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, cognition enhancement, learning and memory enhancement, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism,

hyperparathyroidism, thyroid hormone resistance syndrome, hypercalcemia, hypocalcaemia, hypercholestrolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, blood disorders including hemophilia, dermatological disorders, including, cellulite, acne, eczema, psoriasis and wound healing, scarring, negative effects of aging, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, hormone replacement therapies, steroid hormone-like mediated hair characteristics, immunomodulation, AIDS, vision disorders, glucocorticoid resistance, mineralocorticoid resistance, androgen resistance, pseudohypoaldosteronism, spinal/bulbar muscular atrophy, extraskeletal myxoid chrondrosarcomas, adrenal insufficiency, sexual reversal, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated.

Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, and the LBDG13 polypeptide.

It will be appreciated that modification of detail may be made without departing from the scope of the invention.

Brief description of the Figures

Figure 1: Front page of the Biopendium™. Search initiated using 3ERT:A.

Figure 2A: Inpharmatica Genome Threader™ results of search using 3ERT:A. The arrow points to P06212, Chicken Estrogen Receptor a typical Nuclear Hormone Receptor Ligand Binding Domain family member.

Figure 2B: Inpharmatica Genome Threader™ results of search using 3ERT:A. The arrow points to the CAA05410.2 (LBDG5) protein.

Figure 2C: Inpharmatica PSI-Blast results from search using 3ERT:A. The arrow points to the CAA05410.2 (LBDG5) protein.

Figure 3: InterPro search results for CAA05410.2 (LBDG5).

Figure 4: NCBI Conserved Domain Database search results for CAA05410.2 (LBDG5).

Figure 5A: NCBI protein report for CAA05410.2 (LBDG5).

Figure 5B: SWISS-PROT protein report for O43245, an equivalent sequence to CAA05410.2 (LBDG5).

Figure 5C: Graphical view of NCBI PSI-BLAST iteration 1 to 10 results when queried with CAA05410.2 (LBDG5).

Figure 5D: List of NCBI PSI-BLAST iteration 1 to 10 results when queried with CAA05410.2 (LBDG5).

Figure 6: Inpharmatica Genome Threader™ results of search using CAA05410.2 (LBDG5) as the query sequence. The arrow points to 3ERT:A, the structure of Human Estrogen Receptor alpha Ligand Binding Domain.

Figure 7A: Selection of Inpharmatica reverse-maximised PSI-Blast results from search using CAA05410.2 (LBDG5) as the query sequence. Relationships to known Nuclear Hormone Receptor Ligand Binding Domains are found in the second positive iteration. The arrow points to AAC52143.1, the Nuclear Hormone Receptor Ligand Binding Domain Rat HZF-3.

Figure 7B: Selection of Inpharmatica PSI-Blast reverse-maximised results from search using CAA05410.2 (LBDG5) as the query sequence The arrow points to 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain), the original query structure.

Figure 8: Genome Threader™ alignment of CAA05410.2 (LBDG5) and 3ERT:A. Residues which are conserved between CAA05410.2 (LBDG5) and 3ERT:A in the 3ERT:A dimerisation helix are boxed in black.

Figure 9: RasMol view of 3ERT:A (Human Estrogen Receptor Ligand Binding Domain) with residues in the dimerisation helix which are conserved between CAA05410.2 (LBDG5) and 3ERT:A highlighted in wireframe view of sidechain.

Figure 10: Inpharmatica PSI-BLAST results for CAA05410.2 (LBDG5), arrow 1 points to CAA05409.2 (LBDG10), the rat orthologue of CAA05410.2 (LBDG5). Arrow 2 points to BAB62888.1 (LBDG11), the human paralogue of CAA05410.2 (LBDG5). Arrow 3 points

to AAH03486.1 (LBDG12), the mouse orthologue of BAB62888.1 (LBDG11). Arrow 4 points to AAK49953.1 (LBDG13), the rat orthologue of BAB62888.1 (LBDG11).

Figure 11: Inpharmatica Genome Threader™ results of search using CAA05409.2 (LBDG10) as the query sequence. The arrow points to 3ERT:A, the structure of Human Estrogen Receptor alpha Ligand Binding Domain.

Figure 12: Selection of Inpharmatica PSI-Blast results from search using CAA05409.2 (LBDG10) as the query sequence The arrow points to 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain).

Figure 13: Alignment of CAA05409.2 (LBDG10) with CAA05410.2 (LBDG5) and 3ERT:A. Residues which are conserved between CAA05409.2 (LBDG10), CAA05410.2 (LBDG5) and 3ERT:A in the 3ERT:A dimerisation helix are boxed in black.

Figure 14: InterPro search results for CAA05409.2 (LBDG10).

Figure 15: Conserved Domain (CD) Database search results for CAA05409.2 (LBDG10).

Figure 16A: Graphical view of NCBI PSI-BLAST iteration 1 to 10 results when queried with CAA05409.2 (LBDG10).

Figure 16B: List of NCBI PSI-BLAST iteration 1 to 10 results when queried with CAA05409.2 (LBDG10).

Figure 17: NCBI protein report for CAA05409.2 (LBDG10).

Figure 18: Diagram of relationships between 3ERT:A, CAA05410.2 (LBDG5), BAB62888.1 (LBDG11), AAH03486.1 (LBDG12) and AAK49953.1 (LBDG13).

Figure 19: Genome Threader alignment of 3ERT:A with CAA05410.2 (LBDG5). The sequences of BAB62888.1 (LBDG11), AAH03486.1 (LBDG12) and AAK49953.1 (LBDG13) have been added to the alignment with respect to CAA05410.2 (LBDG5). Grey boxes mark the boundaries of the Genome Threader relationship between 3ERT:A and CAA05410.2 (LBDG5).

Figure 20: InterPro search results for BAB62888.1 (LBDG11).

Figure 21: Conserved Domain (CD) Database search results for BAB62888.1 (LBDG11).

Figure 22: Graphical view of NCBI PSI-BLAST iteration 10 results when queried with BAB62888.1 (LBDG11).

Figure 23: List of NCBI PSI-BLAST iteration 10 results when queried with BAB62888.1 (LBDG11).

Figure 24: NCBI protein report for BAB62888.1 (LBDG11).

Figure 25: InterPro search results for AAH03486.1 (LBDG12).

Figure 26: Conserved Domain (CD) Database search results for AAH03486.1 (LBDG12).

Figure 27: Graphical view of NCBI PSI-BLAST iteration 10 results when queried with AAH03486.1 (LBDG12).

Figure 28: List of NCBI PSI-BLAST iteration 10 results when queried with AAH03486.1 (LBDG12).

Figure 29: NCBI protein report for AAH03486.1 (LBDG12).

Figure 30: InterPro search results for AAK49953.1 (LBDG13).

Figure 31: Conserved Domain (CD) Database search results for AAK49953.1 (LBDG13).

Figure 32: Graphical view of NCBI PSI-BLAST iteration 10 results when queried with AAK49953.1 (LBDG13).

Figure 33: List of NCBI PSI-BLAST iteration 10 results when queried with AAK49953.1 (LBDG13).

Figure 34: NCBI protein report for AAK49953.1 (LBDG13).

Figure 35: Constitutive transcriptional activity of GAL4-LBDG11 LBD

Figure 36: Normalised expression of LBDG11 in 26 samples of 18 normal human tissues.

Figure 37: Normalised expression of LBDG11 in 17 cell line samples.

Figure 38: Normalised expression of LBDG11 in 32 diseased and clinically matched normal samples.

Examples

Example 1: CAA05410.2 (LBDG5)

In order to initiate a search for a novel, distantly related Nuclear Hormone Receptor Ligand Binding Domain, an archetypal Nuclear Hormone Receptor Ligand Binding Domain family member, Human Estrogen Receptor alpha Ligand Binding Domain is chosen. More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is Human Estrogen Receptor alpha Ligand Binding Domain, PDB code 3ERT:A (Figure 1). A search of the Biopendium™ for homologues of 3ERT:A takes place and returns 6310 Inpharmatica Genome Threader™ results (selection given in Figure 2A and 2B) and 1093 Inpharmatica PSI-Blast results (selection in Figure 2C). The 6310 Genome Threader™ results include examples of other Nuclear Hormone Receptor Ligand Binding Domain family members, such as Chicken Estrogen Receptor, marked by an arrow in Figure 2A. Among the known Nuclear Hormone Receptor Ligand Binding Domain members appears a protein of apparently unknown function, CAA05410.2 (LBDG5, marked by an arrow in Figure 2B).

The Inpharmatica Genome Threader™ has identified residues 394-604 of a sequence, CAA05410.2 (LBDG5), as having an equivalent structure to residues 15-247 of Human Estrogen Receptor alpha Ligand Binding Domain (PDB code: 3ERT:A). Having a structure equivalent to 3ERT:A suggests that CAA05410.2 (LBDG5) is a protein that functions as a Nuclear Hormone Receptor Ligand Binding Domain. The Inpharmatica Genome Threader™ identifies this with 100% confidence.

Positive iterations of PSI-Blast are unable to identify this relationship (Forward PSI-Blast does identify known Nuclear Hormone Receptor Ligand Binding Domain members with varying degrees of probability (E-value) as would be expected.). It is only in negative iterations that Inpharmatica PSI-Blast can identify residues 441-518 of CAA05410.2 (LBDG5) as having a sequence relationship to the Human Estrogen Receptor alpha Ligand Binding Domain (Figure 2C). The ability to identify relationships via negative iterations of PSI-Blast is a product of the all-by-all sequence comparison (reverse-maximisation) that underlies the Biopendium and is unique to Inpharmatica. The identification of a relationship

between 3ERT:A and CAA05410.2 (LBDG5) in PSI-Blast iteration –5 at a significant E-value of 2.0E-08 strongly supports the Genome Threader annotation of CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

In order to view the public domain annotation of CAA05410.2 (LBDG5) the InterPro secondary database is queried with CAA05410.2 (LBDG5; Figure 3). It can be seen from Figure 3 that InterPro annotates a region of CAA05410.2 (LBDG5) as containing a C4-type zinc finger. The region of CAA05410.2 (LBDG5) annotated by InterPro as containing a C4-type zinc finger lies between residues 303 and 363. This region is N-terminal to, and does not overlap with, residues 394-604 of CAA05410.2 (LBDG5), which Genome Threader and Inpharmatica PSI-Blast annotate as containing a novel Nuclear Hormone Receptor Ligand Binding Domain. Thus residues 303-363 of CAA05410.2 (LBDG5) constitute a C4-type zinc finger. C4-type zinc fingers are often found N-terminal to Nuclear Hormone Receptor Ligand Binding Domains, and the observation of a C4-type zinc finger N-terminal to the proposed Nuclear Hormone Receptor Ligand Binding Domain further supports the Genome Threader annotation of of CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain. It is important to note, however, that the presence of a C4-type zinc finger does not indicate that a Nuclear Hormone Receptor Ligand Binding Domain must be present. Examples of well-documented proteins which contain a C4-type zinc finger but lack a Nuclear Hormone Receptor Ligand Binding Domain include Knirps (SWISS-PROT accession P10734) and ODR7 (SWISS-PROT accession P41933). Implicit in the existence of proteins such as Knirps and ODR7 is the fact that possession of a DBD does not mean that a LBD will be concomitantly present. It can be seen from Figure 3 that Interpro does not annotate any region of CAA05410.2 as containing a Nuclear Hormone Receptor Ligand Binding Domain. This demonstrates that CAA05410.2 (LBDG5) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using InterPro.

In order to view what is known in the public domain secondary databases, the NCBI Conserved Domain Database (CDD) is queried with CAA05410.2 (LBDG5; Figure 4). CDD returns the C4-type zinc finger as discussed above but does not identify CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain. This

demonstrates that CAA05410.2 (LBDG5) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using CDD.

The National Centre for Biotechnology Information (NCBI) GenBank protein database is viewed to examine if there is any further information that is known in the public domain relating to CAA05410.2 (LBDG5). This is the U.S. public domain database for protein and gene sequence deposition (Figure 5A). CAA05410.2 was cloned by a group of scientists at the Anderson Cancer Centre, University of Texas, USA. The authors annotate CAA05410.2 (LBDG5) as "oncofetal protein p65" and classify it as a steroid/thyroid receptor superfamily member (a synonym for Nuclear Hormone Receptor family). However, this annotation does not reflect identification of a Ligand Binding Domain, instead this annotation of CAA05410.2 (LBDG5) as a Nuclear Hormone Receptor was performed solely on the basis of the "obvious" N-terminal C4-type zinc finger. No public domain annotation was made that CAA05410.2 (LBDG5) contains a Nuclear Hormone Receptor Ligand Binding Domain. Analogous classifications to the Nuclear Hormone receptor family have been made purely of the basis of a protein possessing a C4-type zinc finger (eg. Knirps and ODR7 are referred to as Nuclear Hormone Receptors). Implicit in the existence of proteins such as Knirps and ODR7 is the fact that possession of a C4-type zinc finger DBD does not mean that a LBD will be concomitantly present. Thus the (NCBI) GenBank does not in any way annotate CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain or suggest that the protein has this function.

CAA05410.2 (LBDG5) is also deposited in the SWISS-PROT protein database under the identifier O43245 (Figure 5B). As discussed above, CAA05410.2 (LBDG5) is annotated as belonging to the Nuclear Hormone Receptors family, but there is no annotation that this sequence contains a Ligand Binding Domain. This is confirmed in the database cross-references section of the SWISS-PROT report (arrow Figure 5B) which annotates CAA05410.2 (LBDG5) as containing a C4-type zinc finger, but does not contain any reference or suggestion to the presence of a Nuclear Hormone Receptor Ligand Binding Domain.

NCBI provides a public domain PSI-Blast server. Querying NCBI PSI-Blast with CAA05410.2 (LBDG5) through 10 positive iterations does not provide any obvious relationship of CAA05410.2 (LBDG5)to any known Nuclear Hormone Receptor Ligand

Binding Domains (note that (1) NCBI PSI-Blast has converged by iteration 10, and so further iterations would not return any more sequences and that (2) NCBI PSI-Blast cannot provide data on negative iterations because no all-by-all calculation is performed). Figure 5C shows the graphical display of NCBI PSI-Blast results for CAA05410.2 (LBDG5). The horizontal axis corresponds to N-terminal to C-terminal residue numbering along the CAA05410.2 (LBDG5) protein. It is clear that there are 6 obvious sequences (lines Figure 5C) which are hit by the Nuclear Hormone Receptor Ligand Binding Domain region (residues 394-604) of CAA05410.2 (LBDG5). The accession codes of these 6 sequences in are listed in Figure 5D (marked by lines). None of these 6 sequences have been annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus NCBI PSI-Blast does not annotate residues 394-604 of CAA05410.2 (LBDG5) as having a relationship to any known Nuclear Hormone receptor Ligand Binding Domains in an obvious manner. Note that a large number of Nuclear Hormone Receptor sequences are hit by the region of CAA05410.2 (LBDG5) that contains an "obvious" C4-type zinc finger (in the region of residues 200-400 approximately). This indicates that NCBI PSI-Blast is annotating CAA05410.2 (LBDG5) as having a relationship to known Nuclear Hormone Receptor C4-type zinc fingers, but fails to find any obvious relationships to any known Nuclear Hormone Receptor Ligand Binding Domains.

There is no further public domain annotation for CAA05410.2 (LBDG5). The public domain information for this protein does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain. Therefore using all public domain annotation tools, CAA05410.2 (LBDG5) is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain. Only the Inpharmatica Genome Threader™ and Inpharmatica PSI-Blast are able to annotate this protein as containing a Nuclear Hormone Receptor Ligand Binding Domain.

The reverse search is now carried out. CAA05410.2 (LBDG5) is now used as the query sequence in the Biopendium™. The Inpharmatica Genome Threader™ identifies 157 hits (Figure 6) while Inpharmatica PSI-Blast returns 2093 hits (Figures 7A and 7B). The Inpharmatica Genome Threader™ (Figure 6, arrow) identifies residues 394-604 of CAA05410.2 (LBDG5) as having a structure the same as Human Estrogen Receptor alpha Ligand Binding Domain (PDB code: 3ERT:A) with 100% confidence. Thus a region from

residues 394 to residues 604 of CAA05410.2 (LBDG5) has been identified as adopting an equivalent fold to the Human Estrogen Receptor alpha Ligand Binding Domain.

Inpharmatica PSI-Blast also identifies the same region of CAA05410.2 (LBDG5) as having a relationship with known Nuclear Hormone receptor Ligand Binding Domains by the second positive iteration. For example, Figure 7A shows a selection of Inpharmatica PSI-Blast results and it can be seen that the sequence AAC52143.1 (Rat Nuclear Hormone Receptor HZF-3) has a highly significant relationship to CAA05410.2 (LBDG5), being found in the second positive iteration with an E-value of 1.0E-30. The results indicate that residues 196-599 of CAA05410.2 (LBDG5) are related to residues 146-596 of AAC52143.1 (Rat Nuclear Hormone Receptor HZF-3). Residues 196-599 includes almost all of the CAA05410.2 (LBDG5) Nuclear Hormone Receptor Ligand Binding Domain region identified by Genome Threader™ (residues 394-604), and matches them to a region of HZF-3 which contains a known Nuclear Hormone Receptor Ligand Binding Domain (residues 420-592, as determined by PFAM). Thus Inpharmatica PSI-Blast is in strong agreement with Inpharmatica Genome Threader™ at annotating a region between residues 394 to 604 of CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain. This is in contrast to public domain NCBI PSI-Blast which fails to identify any obvious relationship between CAA05410.2 (LBDG5) and known Nuclear Hormone Receptor Ligand Binding Domains (Figures 5C and 5D). Only Inpharmatica Genome Threader™ and Inpharmatica PSI-Blast are able to identify CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain. Inpharmatica PSI-Blast also identifies a relationship between CAA05410.2 (LBDG5) and the original query structure 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain), Figure 7B arrow. The relationship between CAA05410.2 (LBDG5) and 3ERT:A is found in the fifth positive iteration and has a significant E-value of 2.0E-08. This further consolidates the Genome Threader annotation of CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Among the Nuclear Hormone Receptor Ligand Binding Domain family members that the Inpharmatica Genome Threader™ returns is the original input query Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A). 3ERT:A is chosen against which to view the sequence alignment of CAA05410.2 (the LBDG5 polypeptide). Viewing the

alignment (Figure 8) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology. A particularly interesting feature of the Genome Threader alignment is found in the region that corresponds to the dimerisation helix of 3ERT:A (marked by asterisks in Figure 8). It can be seen that 12 residues of the 3ERT:A dimerisation helix are precisely conserved in CAA05410.2 (LBDG5; marked in black boxes). Particularly striking is the Ser-His-Ile-Arg-His-Met block of 100% identity. Because this helix is the primary interface for Ligand Binding Domain dimerisation, and plays a major role in determining dimer partner specificity, the conservation of residues in the dimerisation helix between 3ERT:A (Human Estrogen Receptor alpha LBD) and CAA05410.2 (LBDG5) strongly suggests that the Ligand Binding Domain of CAA05410.2 (LBDG5) will exhibit dimerisation properties similar to the Human Estrogen Receptor alpha Ligand Binding Domain. Since an Estrogen Receptor alpha Ligand Binding Domain can either homodimerise (ie. dimerise with another Estrogen Receptor alpha Ligand Binding Domain) or heterodimerise with an Estrogen Receptor beta Ligand Binding Domain, this would suggest that the Ligand Binding Domain of CAA05410.2 (LBDG5) will be able to (a) homodimerise with another CAA05410.2 (LBDG5) Ligand Binding Domain, (b) heterodimerise with an Estrogen Receptor alpha Ligand Binding Domain and (c) heterodimerise with an Estrogen Receptor beta Ligand Binding Domain. Predicted heterodimerisation with Estrogen Receptor alpha and Estrogen Receptor beta Ligand Binding Domains implicates CAA05410.2 (LBDG5) in all diseases linked to Estrogen Receptor alpha and Estrogen Receptor beta.

In order to ensure that the protein identified is a homologue of the query sequence, the visualisation program RasMol (Figure 9) is used. This visualization tools identifies important sites of known protein structures. This visualisation is shown with Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), which illustrates the positions of residues on the dimersation helix which are 100% conserved in CAA05410.2 (LBDG5).

To summarise, only Inpharmatica Genome Threader™ can identify that residues 394-604 of CAA05410.2 (LBDG5) folds in a similar manner to 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain) and as such is identified as containing a novel Nuclear Hormone Receptor Ligand Binding Domain. This annotation is also strongly supported by

Inpharmatica PSI-Blast sequence-sequence relationships which relate this region of CAA05410.2 (LBDG5) to known Nuclear Hormone Receptor ligand Binding Domains.

Example 2: CAA05409.2 (LBDG10)

A *Rattus norvegicus* orthologue of CAA05410.2 (LBDG5) has been identified, and will be referred to as CAA05409.2 (LBDG10). CAA05409.2 (LBDG10) is identified in the first iteration of Inpharmatica PSI-BLAST (Figure 10, arrow 1) and has 83% sequence identity with CAA05410.2 (LBDG5). On the basis of the high homology to CAA05410.2 (LBDG5), we annotate CAA05409.2 (LBDG10) as also containing a Nuclear Hormone Receptor Ligand Binding Domain. In addition to this homology, CAA05409.2 (LBDG10) has Genome Threader™ and Inpharmatica PSI-BLAST relationships which consolidate the annotation of CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

CAA05409.2 (LBDG10) is now used as the query sequence in the Biopendium™. The Inpharmatica Genome Threader™ identifies 128 hits (Figure 11) while Inpharmatica PSI-Blast returns 1920 hits (Figure 12). The Inpharmatica Genome Threader™ (Figure 11, arrow) identifies residues 413-603 of CAA05409.2 (LBDG10) as having a structure the same as Human Estrogen Receptor alpha Ligand Binding Domain (PDB code: 3ERT:A) with 100% confidence. Thus a region from residue 413 to residue 603 of CAA05409.2 (LBDG10) has been identified as adopting an equivalent fold to the Human Estrogen Receptor alpha Ligand Binding Domain.

Inpharmatica PSI-Blast also identifies the same region of CAA05409.2 (LBDG10) as having a relationship with known Nuclear Hormone receptor Ligand Binding Domains by the fourth positive iteration. For example, Figure 12 shows a selection of Inpharmatica PSI-Blast results and it can be seen that the sequence of 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain) has a highly significant relationship to CAA05409.2 (LBDG10), being found in the fourth positive iteration with an E-value of $2.0E^{-43}$. The results indicate that residues 413-603 of CAA05409.2 (LBDG10) are related to 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain). Thus Inpharmatica PSI-Blast is in complete agreement with Inpharmatica Genome Threader™ at annotating a region between residues 413 to 603 of CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Among the Nuclear Hormone Receptor Ligand Binding Domain family members that the Inpharmatica Genome Threader™ returns is the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A). 3ERT:A is chosen against which to view the sequence alignment of CAA05409.2 (the LBDG10 polypeptide). CAA05410.2 (LBDG5) is also included in the alignment. Viewing the alignment (Figure 13) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology. A particularly interesting feature of the Genome Threader alignment is found in the region that corresponds to the dimerisation helix of 3ERT:A (marked by asterisks in Figure 13). It can be seen that 12 residues of the 3ERT:A dimerisation helix are precisely conserved in CAA05409.2 (LBDG10; marked in black boxes). Particularly striking is the Ser-His-Ile-Arg-His-Met block of 100% identity. Because this helix is the primary interface for Ligand Binding Domain dimerisation, and plays a major role in determining dimer partner specificity, the conservation of residues in the dimerisation helix between 3ERT:A (Human Estrogen Receptor alpha LBD) and CAA05409.2 (LBDG10) strongly suggests that the Ligand Binding Domain of CAA05409.2 (LBDG10) will exhibit dimerisation properties similar to the Human Estrogen Receptor alpha Ligand Binding Domain. Since an Estrogen Receptor alpha Ligand Binding Domain can either homodimerise (*i.e.* dimerise with another Estrogen Receptor alpha Ligand Binding Domain) or heterodimerise with an Estrogen Receptor beta Ligand Binding Domain, this would suggest that in *Rattus norvegicus* the Ligand Binding Domain of CAA05409.2 (LBDG10) will be able to (a) homodimerise with another CAA05409.2 (LBDG10) Ligand Binding Domain, (b) heterodimerise with a (Rat) Estrogen Receptor alpha Ligand Binding Domain and (c) heterodimerise with a (Rat) Estrogen Receptor beta Ligand Binding Domain. This also suggests that in *Homo sapiens* cell-based experimental systems the Ligand Binding Domain of CAA05409.2 (LBDG10) will be able to (d) heterodimerise with the Human orthologue CAA05410.2 (LBDG5) Ligand Binding Domain, (e) heterodimerise with a (Human) Estrogen Receptor alpha Ligand Binding Domain and (f) heterodimerise with a (Human) Estrogen Receptor beta Ligand Binding Domain. Predicted heterodimerisation with Estrogen Receptor alpha and Estrogen Receptor beta Ligand Binding Domains implicates CAA05409.2 (LBDG10) in all diseases linked to Estrogen Receptor alpha and Estrogen Receptor beta.

In order to view the public domain annotation of CAA05409.2 (LBDG10) the InterPro secondary database is queried with CAA05409.2 (LBDG10; Figure 14). It can be seen from Figure 14 that InterPro annotates a region of CAA05409.2 (LBDG10) as containing a C4-type zinc finger. The region of CAA05409.2 (LBDG10) annotated by InterPro as containing a C4-type zinc finger lies between residues 301 and 369. This region is N-terminal to, and does not overlap with, residues 413-603 of CAA05409.2 (LBDG10), which Genome Threader and Inpharmatica PSI-Blast annotate as containing a novel Nuclear Hormone Receptor Ligand Binding Domain. Thus residues 301-369 of CAA05409.2 (LBDG10) constitute a C4-type zinc finger. C4-type zinc fingers are often found N-terminal to Nuclear Hormone Receptor Ligand Binding Domains, and the observation of a C4-type zinc finger N-terminal to the proposed Nuclear Hormone Receptor Ligand Binding Domain further supports the Genome Threader annotation of of CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain. It is important to note, however, that the presence of a C4-type zinc finger does not indicate that a Nuclear Hormone Receptor Ligand Binding Domain must be present. Examples of well-documented proteins which contain a C4-type zinc finger but lack a Nuclear Hormone Receptor Ligand Binding Domain include Knirps (SWISS-PROT accession P10734) and ODR7 (SWISS-PROT accession P41933). Implicit in the existence of proteins such as Knirps and ODR7 is the fact that possession of a DBD does not mean that a LBD will be concomitantly present. It can be seen from Figure 14 that Interpro does not annotate any region of CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain. This demonstrates that CAA05409.2 (LBDG10) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using InterPro.

In order to view what is known in the public domain secondary databases, the NCBI Conserved Domain Database (CDD) is queried with CAA05409.2 (LBDG10; Figure 15). CDD returns the C4-type zinc finger as discussed above but does not identify CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain. This demonstrates that CAA05409.2 (LBDG10) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using CDD.

NCBI provides a public domain PSI-Blast server. Querying NCBI PSI-Blast with CAA05409.2 (LBDG10) through 10 positive iterations does not provide any obvious

relationship of CAA05409.2 (LBDG10) to any known Nuclear Hormone Receptor Ligand Binding Domains (note that (1) NCBI PSI-Blast has converged by iteration 10, and so further iterations would not return any more sequences and that (2) NCBI PSI-Blast cannot provide data on negative iterations because no all-by-all calculation is performed). Figure 16A shows the graphical display of NCBI PSI-Blast results for CAA05409.2 (LBDG10). The horizontal axis corresponds to N-terminal to C-terminal residue numbering along the CAA05409.2 (LBDG10) protein. It is clear that there are 6 obvious sequences (lines Figure 16A) which are hit by the Nuclear Hormone Receptor Ligand Binding Domain region (residues 413-603) of CAA05409.2 (LBDG10). The accession codes of these 6 sequences in are listed in Figure 16B (marked by lines). None of these 6 sequences have been annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus NCBI PSI-Blast does not annotate residues 413-603 of CAA05409.2 (LBDG10) as having a relationship to any known Nuclear Hormone receptor Ligand Binding Domains in an obvious manner. Note that a large number of Nuclear Hormone Receptor sequences are hit by the region of CAA05409.2 (LBDG10) that contains an "obvious" C4-type zinc finger (in the region of residues 200-400 approximately). This indicates that NCBI PSI-Blast is annotating CAA05409.2 (LBDG10) as having a relationship to known Nuclear Hormone Receptor C4-type zinc fingers, but fails to find any obvious relationships to any known Nuclear Hormone Receptor Ligand Binding Domains.

The National Centre for Biotechnology Information (NCBI) GenBank protein database is viewed to examine if there is any further information that is known in the public domain relating to CAA05409.2 (LBDG10). This is the U.S. public domain database for protein and gene sequence deposition (Figure 17). CAA05409.1 was cloned by a group of scientists at the Anderson Cancer Centre, University of Texas, USA. The authors annotate CAA05409.2 (LBDG10) as "oncofetal protein p65" and classify it as a steroid/thyroid receptor superfamily member (a synonym for Nuclear Hormone Receptor family). However, this annotation does not reflect identification of a Ligand Binding Domain, instead this annotation of CAA05409.2 (LBDG10) as a Nuclear Hormone Receptor was performed solely on the basis of the "obvious" N-terminal C4-type zinc finger. No public domain annotation was made that CAA05409.2 (LBDG10) contains a Nuclear Hormone Receptor Ligand Binding Domain. Analogous classifications to the

Nuclear Hormone receptor family have been made purely of the basis of a protein possessing a C4-type zinc finger (eg. Knirps and ODR7 are referred to as Nuclear Hormone Receptors). Implicit in the existence of proteins such as Knirps and ODR7 is the fact that possession of a C4-type zinc finger DBD does not mean that a LBD will be concomitantly present. Thus the (NCBI) GenBank does not in any way annotate CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain or suggest that the protein has this function.

There is no further public domain annotation for CAA05409.2 (LBDG10). The public domain information for this protein does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain. Therefore using all public domain annotation tools, CAA05409.2 (LBDG10) is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain.

To summarise, Inpharmatica Genome Threader™ can identify that residues 413-603 of CAA05409.2 (LBDG10) folds in a similar manner to 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain) and as such is identified as containing a novel Nuclear Hormone Receptor Ligand Binding Domain. This annotation is also strongly supported by Inpharmatica PSI-Blast sequence-sequence relationships which relate this region of CAA05409.2 (LBDG10) to known Nuclear Hormone Receptor ligand Binding Domains. This is further consolidated by the observation that CAA05409.2 (LBDG10) has high sequence homology to CAA05410.2 (LBDG5).

Example 3: BAB62888.1 (LBDG11)

A *Homo sapiens* paralogue of CAA05410.2 (LBDG5) has been identified, and will be referred to herein as BAB62888.1 (LBDG11). Residues 52-329 of BAB62888.1 (LBDG11) are identified in the first iteration of Inpharmatica PSI-BLAST (Figure 10, arrow 2) as sharing 60% sequence identity with residues 330-614 of CAA05410.2 (LBDG5). Residues 330-614 of CAA05410.2 (LBDG5) contain the region (residues 394-604) predicted to adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain. On the basis of the high sequence identity that BAB62888.1 (LBDG11) shares with the region of CAA05410.2 (LBDG5) that is predicted to adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain, we predict that BAB62888.1 (LBDG11) also adopts the structure of a Nuclear Hormone Receptor Ligand Binding Domain. Chothia and Lesk, 1986 (EMBO

Journal vol.5 pp823) first showed that for proteins with more than 50% sequence identity 85% of residues would adopt the same conformation. Other groups (Sander, C. and Schneider, R. (1991) Proteins vol.9 pp56; Hubbard, T.J.P. and Blundell, T.L. (1987) Protein Engineering vol.1 pp159; Flores, T.P., Orengo, C.A., Moss, D.M. and Thornton, J.M. (1993) Protein Science vol.2 pp1811; and Hilbert, M., Bohm, G. and Jaenicke, R. (1993) Proteins vol.17 pp138) subsequently extended these studies and have showed that the fold remains the same even if sequence identity falls as low as 30%.

Figure 18 schematically depicts the relationships between the structure of the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), CAA05410.2 (LBDG5) and BAB62888.1 (LBDG11). Genome Threader™ and Inpharmatica PSI-BLAST have identified residues 394-604 of CAA05410.2 (LBDG5) as adopting the structure of residues 15-247 (residues 320-552 in full-length numbering scheme) of the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), double-headed arrow at the top of Figure 18. The relationships between 3ERT:A, CAA05410.2 (LBDG5) and BAB62888.1 (LBDG11) are also presented in the form of a multiple alignment (Figure19). This alignment has been generated by first taking the Genome Threader alignment of 3ERT:A with CAA05410.2 (LBDG5), and then adding the sequence of BAB62888.1 (LBDG11) to the alignment on the basis of it's sequence identity to CAA05410.2 (LBDG5). By reference to the alignment it is possible to map the domain boundaries of the predicted Ligand Binding Domain of CAA05410.2 (LBDG5) onto BAB62888.1 (LBDG11): these domain boundary residues are marked by grey boxes in Figure 19. This identifies that residues 118-319 of BAB62888.1 (LBDG11) correspond to the residues 394-604 of CAA05410.2 (LBDG5) that are predicted to adopt the structure of a Ligand Binding Domain by Genome Threader and Inpharmatica PSI-BLAST (see also Figure 18). Thus we annotate residues 118-319 of BAB62888.1 (LBDG11) as adopting the structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5).

In order to view the public domain annotation of BAB62888.1 (LBDG11) the InterPro secondary database is queried with BAB62888.1 (LBDG11; Figure 20). It can be seen from Figure 20 that no matches are found to InterPro. Returning no matches demonstrates that

BAB62888.1 (LBDG11) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using InterPro.

In order to view what is known in the public domain secondary databases, the NCBI Conserved Domain Database (CDD) is queried with BAB62888.1 (LBDG11; Figure 21). CDD returns no hits. Returning no hits demonstrates that BAB62888.1 (LBDG11) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using CDD.

NCBI provides a public domain PSI-Blast server. Querying NCBI PSI-Blast with BAB62888.1 (LBDG11) through 10 positive iterations fails to annotate any region of BAB62888.1 (LBDG11) as having a relationship to any known Nuclear Hormone Receptor Ligand Binding Domains (note that NCBI PSI-Blast cannot provide data on negative iterations because no all-by-all calculation is performed). Figure 22 shows the graphical display of NCBI PSI-Blast results for BAB62888.1 (LBDG11). Figure 23 lists these NCBI PSI-BLAST hits (marked by lines). None of these sequences have been annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus NCBI PSI-Blast does not annotate BAB62888.1 (LBDG11) as having a relationship to any known Nuclear Hormone receptor Ligand Binding Domains.

The National Centre for Biotechnology Information (NCBI) GenBank protein database is viewed to examine if there is any further information that is known in the public domain relating to BAB62888.1 (LBDG11). This is the U.S. public domain database for protein and gene sequence deposition (Figure 24). BAB62888.1 was cloned by a group of scientists at the Science University of Tokyo, Japan. The authors annotate BAB62888.1 (LBDG11) as a protein that interacts with Terminal Deoxynucleotidyl Transferase (TdT). TdT is a DNA polymerase that enhances Ig and TcR gene diversity in the N region in B- and T-cells. Annotation as a TdT interacting protein has no connection to the Nuclear Hormone Receptor Ligand Binding Domain superfamily. However, informed by the Inpharmatica annotation of BAB62888.1 (LBDG11) as containing a novel Nuclear Hormone Receptor Ligand Binding Domain we are able to predict that BAB62888.1 (LBDG11) will regulate TdT activity in response to steroid-like ligands. Thus the (NCBI) GenBank does not in any way annotate BAB62888.1 (LBDG11) as containing a Nuclear Hormone Receptor Ligand Binding Domain or suggest that the protein has this function.

There is no further public domain annotation for BAB62888.1 (LBDG11). The public domain information for this protein does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain. Therefore using all public domain annotation tools, BAB62888.1 (LBDG11) is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain. To summarise we annotate residues 118-319 of BAB62888.1 (LBDG11) as adopting the structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5) which Genome Threader and Inpharmatica PSI-BLAST predict will adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain.

Cloning of the LBD of LBDG11

A. cDNA source for PCR of the LBD

500ng total RNA (Ambion Europe, UK or Clontech Europe, Belgium) from different human tissues was used to generate cDNA using the Superscript RT (Invitrogen) and oligo dT primer following the manufacturer's protocol. 2µl of the reaction were used in the PCR.

B. Cloning of the LBD

Primers LBDG11 F and LBDG11 R (as shown below) were used to amplify the proposed LBD encompassing amino acids 100-329 of LBDG11 from human cDNA synthesized from RNA prepared from stomach, testis, ovary, placenta and brain (as the skilled person appreciates, for practical purposes of cloning, domains are cloned with flanking sequences, hence the region 100-329 was cloned rather than 118-319). PCR was carried out using the DyNAzyme EXT DNA Polymerase (Finnzymes Oy, Espoo, Finland) in 2 mM $MgCl_2$. The resulting PCR product from stomach cDNA was then cloned into the vector pGEMTEasy (Promega UK Ltd, Southampton, UK) and verified by sequence analysis. Sequences were identical to the published sequence of LBDG11. Inserts were then cloned into the vector pFA-CMV (Stratagene Europe, Amsterdam, The Netherlands) by restriction digest with the enzymes BamHI and HindIII. This plasmid expresses the LBD of LBDG11 in fusion with the GAL4 DNA binding domain.

LBDG11 F CCG GGA TCC GTG CGA GAC AAT GTT GGG GA

LBDG11 R CAG AAG CTT TAT CAG GTC TGT GGA GGT G

Identification of NR agonists or antagonists in a cellular assay

A. Assay principle and constructs

The pFA-CMV LBDG11 LBD construct was cotransfected into mammalian cells together with a reporter plasmid, pFR-Luc (Stratagene Europe, Amsterdam, The Netherlands) which contains a synthetic promoter with five tandem repeats of the GAL4 binding sites that control the expression of the luciferase gene. When a potential NR ligand is added to the cells, it will bind the GAL4-LBD and induce transcription of the luciferase gene. The level of the luciferase expression can be monitored by its activity using a luminescence reader (Lehman *et al.* JBC 270, 12953, 1995; Pawar *et al.* JBC, 277, 39243, 2002).

For measurement of constitutive activity, HEK 293 cells were transiently transfected with the reporter pFR-Luc in the absence or presence of the GAL4-LBDG11 LBD or a control GAL4 construct. Luciferase expression was quantified 48h following the transfection.

B. Choosing a cell line and the transfection method

Initially, the cotransfection of the GAL4 construct and reporter constructs has been carried out in HEK 293 cell line. Subsequently, the assay was repeated in a different cell line in which the NR endogenous expression was highest.

The GAL4-LBD and the reporter construct were transiently transfected using the Fugene reagent (Roche) and conditions optimised for each cell line.

C. Compound selection and assay conditions

Various compound libraries were screened in the cellular assay at 10μM final concentration and the activity of potential activators has been confirmed by EC50 determination.

D. Results

A number of potential activators including steroids have been identified in the GAL4 cellular assay.

Identification of ligand binding domain agonists using an *in vitro* assay

This method uses a time resolved fluorescence (TRF) assay to detect ligand-dependent NHR interaction with coactivators.

A critical step in NHR action is the ligand-dependent recruitment of transcriptional coactivators to target gene promoters. Specifically it has been shown that agonist binding to the receptor induces a conformational change that permits the formation of a hydrophobic pocket enabling the receptor to interact with the LXXLL motif (where L is leucine and X any amino acid) contained in most coactivators. Building on this observation, an *in vitro* assay has been designed to screen for putative ligands that will bind the novel NHR in the presence of a peptide containing the LXXLL motif. Biotinylated peptides based on known coactivator sequences, and control peptides (for example a peptide which interacts specifically with ERa) have been synthesized. As an example, the SRC-1 based peptide:

(BIOT-CPSSHSHRLSLTERHKILHRLLQEGSPS-CONH2)

binds the majority of known NHRs, and therefore is expected to interact with the novel NHRs (Mol. Endocrinol. 14, 2010; 2000). TRF assay is performed in solution, by incubating a GST fusion protein containing the LBD of the NHR of interest, europium-labelled anti GST antibody, biotinylated LXXLL containing peptide and streptavidin APC. In the presence of a ligand that binds the LBD, the fluorescent reagents are brought in close proximity and the TRF can occur. A fluorescent signal, which reflects the extent of binding, is read at 665 nm (J. Biol. Mol. Screen. 7, 3; 2002).

Identification of a DNA motif recognised by the DNA binding domain

The DNA Binding Domains (DBDs) of NHR make specific base-specific contacts with the nucleotide bases in the DNA and are responsible for the sequence-specificity of DNA recognition. The NHR response elements contain similar motifs frequently arranged as direct or inverted repeats with a variable spacer. The specificity of NHRs DNA recognition motifs is further enhanced by the formation of homodimers or heterodimers with different partners (e.g. retinoic acid receptor, RXR).

We designed an in vitro assay to rapidly identify the DNA binding motif recognized by the novel NHR, both as a homodimer or heterodimer. The assay is based on an ELISA method carried out in 96 well plates. This approach allows the screening of a large number of putative DNA response elements. Biotinylated oligonucleotides based on a number of binding elements were generated, annealed to form a double-stranded DNA molecule and bound to the 96-well streptavidin coated plates. Excess unbound DNA was washed twice using washing buffer (PBS, 0.1% Tween 20) and following blocking with 2% BSA in PBS, purified recombinant NHR either alone or in the presence of a putative partner protein was added to the plates in binding buffer (PBS, 0.1% Tween 20, 0.1% BSA, 0.1 mg/ml salmon sperm DNA). The bound receptor was detected using general ELISA methodology (Ed Harlow and David Lane, Antibodies – A laboratory manual, Cold Spring Harbor, 2001) by a specific antibody, followed by a secondary HRP conjugated antibody.

Quantitative Expression Profiling of LBDG11

In order to determine the expression profile of the proposed LBDG11, Taqman RT-PCR quantitation was used. The TaqMan 3'- 5' exonuclease assay signals the formation of PCR amplicons by a process involving the nucleolytic degradation of a doublelabeled fluorogenic probe that hybridises to the target template at a site between the two primer recognition sequences (cf. U. S. Patent 5,876,930). The MJ Research Opticon and ABI Prism 7700 automates the detection and quantitative measurement of these signals, which are stoichiometrically related to the quantities of amplicons produced, during each cycle of amplification. In addition to providing substantial reductions in the time and labour requirements for PCR analyses, this technology permits simplified and potentially highly accurate quantification of target sequences in the reactions.

A. RNA samples

Human RNA prepared from diseased or non-diseased organs was purchased from either Ambion Europe (Huntingdon, UK), Clontech (BD, Franklin Lakes, NJ), Biochain (AMS Biotechnology (Europe) Ltd, Abingdon, UK) or Clinomics (Pittsfield, MA).

Cell lines were passaged using standard tissue culture protocols. Cells were maintained in a humidified atmosphere at 37°C; 5% CO_2. Culture medium consisted of 1x DMEM

(Invitrogen, UK) with the addition of 10% foetal bovine serum (Invitrogen, UK); 2mM glutamine (Sigma, Poole, UK); 100 u/ml penicillin G (Invitrogen, UK) and 100u/ml streptomycin sulfate (Invitrogen, UK). Cell treatments were carried out as follows:

Interferon-γ: 200ng/ml for 6 hours

Dibutyryl cAMP: 1mM for 24 hours

PMA: 1μM for 24 hours

All compounds were purchased from Sigma (Poole, UK). RNA was prepared from cell lines using RNeasy kits following manufacturer's protocols (Qiagen, UK)

B. Oligo Design

Oligonucleotide primers and probes were designed using Primer Express software (Applied Biosystems, Foster City CA) with a GC-content of 40-60%, no G-nucleotide at the 5'-end of the probe, and no more than 4 contiguous Gs. Primer probe sets were designed within the 3'exon of the proposed LBDG11 and also over an exon-exon junction of the proposed LBDG11.

Each primer and probe was analysed using BLAST® (Basic Local Alignment Search Tool, Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.: J Mol Biol 1990 Oct 5;215(3):403-10). Results confirmed that each oligonucleotide recognised the target sequence with a specificity >3 bp when compared to other known cDNAs or genomic sequence represented in the Unigene and GoldenPath publicly available databases.

The sequence of the primers and probes used are:

LBDG11 3' Fwd AGCTAGAGGAATTGAAATCCTTTGTC

LBDG11 3' Probe TCCATATACTTTCTCATCTTCTCCACCATCCAGG

LBDG11 3' Rev GATGCTCATTCTCTGTCCGTAGTG

LBDG11 exon Fwd CTGAATGAATCTACCACCTTTGTGTT

LBDG11 exon Probe TCTCGAGCCAACAAAGCCCTGGG

LBDG11 exon Rev TGTGGGTGCTTGATGTAGATTCTT

18s pre-optimised primers and probe were purchased from Applied Biosystems, Foster City, CA.

Probes are covalently conjugated with a fluorescent reporter dye (*e.g.* 6carboxy-fluorescein [FAM]; Xem = 518nm) and a fluorescent quencher dye (6carboxytetram-ethyl-rhodamine [TAMRA]; Mem = 582nm) at the most 5' and most 3' base, respectively. Primers are obtained from Sigma Genosys, UK and probes are obtained from Eurogentec, Belgium.

Primer/probe concentrations were titrated in the range of 50nM to 900nM and optimal concentrations for efficient PCR reactions are determined. Optimal primer and probe concentrations varied in between 100nM and 900nM depending on the target gene that is amplified.

C. cDNA reaction

cDNA was prepared using components from Applied Biosystems, Foster City CA. 50µl reactions are prepared in 0.5ml RNase free tubes. Reactions contained 500ng total RNA; 1x reverse transcriptase buffer; 5.5mM MgCl2; 1mM dNTP's; 2.5µl random hexamers; 20U RNase inhibitor; and 62.5U reverse transcriptase.

D. PCR reactions

25µl reactions were prepared in 0.5 ml thin-walled, optical grade PCR 96 well plates (Applied Biosystems, Foster City CA). Reactions contain: 1x final concentration of TaqMan Universal Master Mix (a proprietary mixture of AmpliTaq Gold DNA polymerase, AmpEraseX UNG, dNTPs with UTP, passive reference dye and optimised buffer components, Applied Biosystems, Foster City CA); 100nM Taqman probe; 900nM forward primer; 900nM reverse primer and the relevant amount of cDNA template. For the normal and diseased tissue analysis, reactions contain the same amounts of reagents except for the probe, which is used at a final concentration of 200nM.

E. Performance of Assay

Standard procedures for the operation of the ABI Prism 7000 or similar detection system were used. This included, for example with the ABI Prism 7000, use of all default program settings with the exception of reaction volume which is changed from 50 to 25 ul. Thermal cycling conditions consisted of two min at 50°C, 10 min at 95°C, followed by

40 cycles of 15 sec at 95°C and 1 min at 60°C. Cycle threshold (Ct) determinations, *i.e.* non-integer calculations of the number of cycles required for reporter dye fluorescence resulting from the synthesis of PCR products to become significantly higher than background fluorescence levels are automatically performed by the instrument for each reaction using default parameters. Assays for target sequences and ribosomal 18s (reference) sequences in the same cDNA samples were performed in separate reaction tubes.

Within each experiment, a standard curve was produced for a typical tissue sample. From this standard curve, the amount of actual starting target or reference cDNA in each test sample was determined.

The levels of target cDNA in each sample were normalised to the level of expression of target in a comparative sample. The levels of internal control cDNA in each sample were normalised to the level of expression of internal control in a comparative sample. The data was then represented as fold expression of normalised target sequence relative to the level of expression in the comparative sample, which was set arbitrarily to 1.

Results

A. Transcriptional Activity of GAL4-LBDG11 LBD

When transfected into cells the GAL4-LBDG11 LBD construct possessed a strong transcriptional activity towards the reporter gene in the absence of an exogenous ligand (Figure 35). This is different from the classical nuclear receptors, which require a ligand for their transcriptional activity. However, similar constitutive activity was observed in the case of several orphan nuclear receptors. For example, the constitutive androstane receptor (CAR) has a high level of activity in a ligand-independent manner due to novel intra-molecular interactions favoured by unique structural features (Dussault *et al.* Mol. Cell Biol. 22, 5270, 2002). A different mechanism seems to be responsible for the constitutive activity of another nuclear receptor, HNF4. The crystal structure of this receptor revealed the constitutive presence of a fatty acid within the ligand binding domain pocket (Wisely *et al.*, Structure 10, 1225, 2002).

B. Expression in Normal Human Tissues

Taqman RT-PCR was carried out using 15ng of the indicated cDNA using primers/probes specific for LBDG11 and 18s rRNA as described in the detailed description. A standard curve for target and internal control was also carried out, using between 25ng to 1.56ng of cDNA template of a typical tissue sample. Using linear regression analysis of the standard curves, the amount of actual starting target or 18s cDNA in each test sample was calculated.

The levels of target and reference cDNA in each sample were normalised to the level of expression in a comparative sample, in this case, stomach. Figure 36 represents the fold expression of normalised target sequence relative to the level of expression in stomach cDNA, which is set arbitrarily to 1. The target expression profile was determined using 2 primer probe sets – one within the 3'exon of the proposed LBDG11 and one over an exon-exon junction within the LBD of LBDG11. Each sample was quantitated in 2 or more individual experiments. Figure 36 shows the mean ±SEM for the multiple experiments.

LBDG11 is expressed at very low levels in most tissues. In general, expression is relatively invariant across the tissues. The relative levels of expression across the different tissues is very similar for both primer/probe sets (3'utr and exon spanning), indicating the absence of differential splicing of LBDG11 in these samples. The PCR reaction was also carried out on cDNA made in the absence of reverse transcriptase enzyme. A signal was not seen in these reactions (data not shown), indicating that the levels of LBDG11 detected are present in cDNA.

C. Expression in Cell Line Samples

Taqman RT-PCR was carried out using 15ng of the indicated cDNA using primers/probes specific for LBDG11 and 18s rRNA as described in the detailed description. A standard curve for target and internal control was also carried out, using between 25ng to 1.56ng of cDNA template of a typical cell line sample. Using linear regression analysis of the standard curves, the amount of actual starting target or 18s cDNA in each test sample was calculated.

The levels of target and reference cDNA in each sample were normalised to the level of expression of target in a comparative sample, in this case, HEK293 cells. Figure 37

represents the fold expression of normalised target sequence relative to the level of expression in HEK293 cells, which is set arbitrarily to 1. The target expression profile was determined using 2 primer probe sets – one within the 3'exon of the proposed LBDG11 and one over an exon-exon junction within the LBD of LBDG11. Each sample was quantitated in 2 or more individual experiments. Figure 37 shows the mean ±SEM for the multiple experiments.

LBDG11 is expressed at very low levels in all cell types, although HL60 appear to have higher relative levels than other cells. In U937 cells, LBDG11 expression is not regulated by any treatment. The relative levels of expression across the different cell types is very similar for both primer/probe sets (3'exon and exon spanning), indicating the absence of differential splicing of LBDG11 in these samples. The PCR reaction was also carried out on cDNA made in the absence of reverse transcriptase enzyme. A signal was not seen in these reactions (data not shown), indicating that the levels of LBDG11 detected are present in cDNA.

Taqman RT-PCR was carried out using 25ng of the indicated cDNA using primers/probes specific for LBDG11 and cyclophilin mRNA as described in the detailed description. A standard curve for target and internal control was also carried out, using between 50ng to 0.78ng of cDNA template of a typical sample. Using linear regression analysis of the standard curves, the amount of actual starting target or cyclophilin cDNA in each test sample was calculated.

The levels of target and reference cDNA in each sample were normalised to the level of expression of target in a comparative sample, in this case, lung cancer normal. Figure 37 represents the fold expression of normalised target sequence relative to the level of expression in lung cancer normal, which is set arbitrarily to 1. The target expression profile was determined using the primer probe set within the 3'utr of the proposed LBDG11. Each sample was quantitated in 2 individual experiments. Figure 38 shows the mean of 2 wells analysed in a representative experiment.

LBDG11 expression is detected in all tissues tested. Expression is low in the Alzheimer's disease, multiple sclerosis and diabetic adipose samples. The highest expression is detected in the diabetic skeletal muscle samples. In Figure 38, it can be seen that LBDG11 is up-regulated 2 fold in prostate cancer when compared to the normal sample; a smaller increase

in expression was observed in breast cancer compared to control. Expression was down-regulated 2-fold in cervical cancer compared to normal control.

For these samples, the PCR reaction was also carried out on cDNA made in the absence of reverse transcriptase enzyme. No signal was seen in these reactions (data not shown), indicating that the levels of LBDG11 detected are present in cDNA.

Example 4: AAH03486.1 (LBDG12)

A *Mus musculus* orthologue of BAB62888.1 (LBDG11) has been identified and will be referred to herein as AAH03486.1 (LBDG12). Residues 52-328 of AAH03486.1 (LBDG12) are identified in the first iteration of Inpharmatica PSI-BLAST (Figure 10, arrow 3) as sharing 62% sequence identity with residues 330-614 of CAA05410.2 (LBDG5). Residues 330-614 of CAA05410.2 (LBDG5) contain the region (residues 394-604) predicted to adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain. On the basis of the high sequence identity that AAH03486.1 (LBDG12) shares with the region of CAA05410.2 (LBDG5) that is predicted to adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain, we predict that AAH03486.1 (LBDG12) also adopts the structure of a Nuclear Hormone Receptor Ligand Binding Domain. Chothia and Lesk, 1986 (EMBO Journal vol.5 pp823) first showed that for proteins with more than 50% sequence identity 85% of residues would adopt the same conformation. Other groups (Sander, C. and Schneider, R. (1991) Proteins vol.9 pp56; Hubbard, T.J.P. and Blundell, T.L. (1987) Protein Engineering vol.1 pp159; Flores, T.P., Orengo, C.A., Moss, D.M. and Thornton, J.M. (1993) Protein Science vol.2 pp1811; and Hilbert, M., Bohm, G. and Jaenicke, R. (1993) Proteins vol.17 pp138) subsequently extended these studies and have showed that the fold remains the same even if sequence identity falls as low as 30%.

Figure 18 schematically depicts the relationships between the structure of the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), CAA05410.2 (LBDG5) and AAH03486.1 (LBDG12). Genome Threader™ and Inpharmatica PSI-BLAST have identified residues 394-604 of CAA05410.2 (LBDG5) as adopting the structure of residues 15-247 of the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), double-headed arrow at the top of Figure 18. The relationships between (3ERT:A), CAA05410.2 (LBDG5) and AAH03486.1 (LBDG12) are also presented in the form of a multiple alignment (Figure19). This alignment has been generated by first taking the Genome

Threader alignment of 3ERT:A with CAA05410.2 (LBDG5), and then adding the sequence of AAH03486.1 (LBDG12) to the alignment on the basis of it's sequence identity to CAA05410.2 (LBDG5). By reference to the alignment it is possible to map the domain boundaries of the predicted Ligand Binding Domain of CAA05410.2 (LBDG5) onto AAH03486.1 (LBDG12), these domain boundary residues are marked by grey boxes in Figure 19. This identifies that residues 118-318 of AAH03486.1 (LBDG12) correspond to the residues 394-604 of CAA05410.2 (LBDG5) that are predicted to adopt the structure of a Ligand Binding Domain by Genome Threader and Inpharmatica PSI-BLAST (see also Figure 18). Thus we annotate residues 118-318 of AAH03486.1 (LBDG12) as adopting the structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5).

In order to view the public domain annotation of AAH03486.1 (LBDG12) the InterPro secondary database is queried with AAH03486.1 (LBDG12; Figure 25). It can be seen from Figure 25 that no matches are found to InterPro. Returning no matches demonstrates that AAH03486.1 (LBDG12) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using InterPro.

In order to view what is known in the public domain secondary databases, the NCBI Conserved Domain Database (CDD) is queried with AAH03486.1 (LBDG12; Figure 26). CDD returns no hits. Returning no hits demonstrates that AAH03486.1 (LBDG12) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using CDD.

NCBI provides a public domain PSI-Blast server. Querying NCBI PSI-Blast with AAH03486.1 (LBDG12) through 10 positive iterations fails to annotate any region of AAH03486.1 (LBDG12) as having a relationship to any known Nuclear Hormone Receptor Ligand Binding Domains (note that NCBI PSI-Blast cannot provide data on negative iterations because no all-by-all calculation is performed). Figure 27 shows the graphical display of NCBI PSI-Blast results for AAH03486.1 (LBDG12). Figure 28 lists these NCBI PSI-BLAST hits (marked by lines). None of these sequences have been annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus NCBI PSI-Blast does not annotate AAH03486.1 (LBDG12) as having a relationship to any known Nuclear Hormone receptor Ligand Binding Domains.

The National Centre for Biotechnology Information (NCBI) GenBank protein database is viewed to examine if there is any further information that is known in the public domain relating to AAH03486.1 (LBDG12). This is the U.S. public domain database for protein and gene sequence deposition (Figure 29). AAH03486.1 was cloned by the I.M.A.G.E. Consortium. AAH03486.1 is annotated as being similar to p65 protein, but as discussed earlier, p65 (CAA05410.2 (LBDG5)) is not annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus the (NCBI) GenBank does not in any way annotate AAH03486.1 (LBDG12) as containing a Nuclear Hormone Receptor Ligand Binding Domain or suggest that the protein has this function.

There is no further public domain annotation for AAH03486.1 (LBDG12). The public domain information for this protein does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain. Therefore using all public domain annotation tools, AAH03486.1 (LBDG12) is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain.

To summarise we annotate residues 118-318 of AAH03486.1 (LBDG12) as adopting the structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5) which Genome Threader and Inpharmatica PSI-BLAST predict will adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain.

Example 5: AAK49953.1 (LBDG13)

A *Rattus norvegicus* orthologue of BAB62888.1 (LBDG11) has been identified and will be referred to herein as AAK49953.1 (LBDG13). Residues 51-327 of AAK49953.1 (LBDG13) are identified in the first iteration of Inpharmatica PSI-BLAST (Figure 10, arrow 4) as sharing 59% sequence identity with residues 330-614 of CAA05410.2 (LBDG5). Residues 330-614 of CAA05410.2 (LBDG5) contain the region (residues 394-604) predicted to adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain. On the basis of the high sequence identity that AAK49953.1 (LBDG13) shares with the region of CAA05410.2 (LBDG5) that is predicted to adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain, we predict that AAK49953.1 (LBDG13) also adopts the structure of a Nuclear Hormone Receptor Ligand Binding Domain. Chothia and Lesk, 1986 (EMBO Journal vol.5 pp823) first showed that for proteins with more than 50% sequence identity

85% of residues would adopt the same conformation. Other groups (Sander, C. and Schneider, R. (1991) Proteins vol.9 pp56; Hubbard, T.J.P. and Blundell, T.L. (1987) Protein Engineering vol.1 pp159; Flores, T.P., Orengo, C.A., Moss, D.M. and Thornton, J.M. (1993) Protein Science vol.2 pp1811; and Hilbert, M., Bohm, G. and Jaenicke, R. (1993) Proteins vol.17 pp138) subsequently extended these studies and have showed that the fold remains the same even if sequence identity falls as low as 30%.

Figure 18 schematically depicts the relationships between the structure of the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), CAA05410.2 (LBDG5) and AAK49953.1 (LBDG13). Genome Threader™ and Inpharmatica PSI-BLAST have identified residues 394-604 of CAA05410.2 (LBDG5) as adopting the structure of residues 15-247 of the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), double-headed arrow at the top of Figure 18. The relationships between (3ERT:A), CAA05410.2 (LBDG5) and AAK49953.1 (LBDG13) are also presented in the form of a multiple alignment (Figure19). This alignment has been generated by first taking the Genome Threader alignment of 3ERT:A with CAA05410.2 (LBDG5), and then adding the sequence of AAK49953.1 (LBDG13) to the alignment on the basis of it's sequence identity to CAA05410.2 (LBDG5). By reference to the alignment it is possible to map the domain boundaries of the predicted Ligand Binding Domain of CAA05410.2 (LBDG5) onto AAK49953.1 (LBDG13), these domain boundary residues are marked by grey boxes in Figure 19. This identifies that residues 117-317 of AAK49953.1 (LBDG13) correspond to the residues 394-604 of CAA05410.2 (LBDG5) that are predicted to adopt the structure of a Ligand Binding Domain by Genome Threader and Inpharmatica PSI-BLAST (see also Figure 18). Thus we annotate residues 117-317 of AAK49953.1 (LBDG13) as adopting the structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5).

In order to view the public domain annotation of AAK49953.1 (LBDG13) the InterPro secondary database is queried with AAK49953.1 (LBDG13; Figure 30). It can be seen from Figure 30 that no matches are found to InterPro. Returning no matches demonstrates that AAK49953.1 (LBDG13) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using InterPro.

In order to view what is known in the public domain secondary databases, the NCBI Conserved Domain Database (CDD) is queried with AAK49953.1 (LBDG13; Figure 31). CDD returns only 1hit to the PFAM profile PF02562, PhoH. This profile match may be a false positive since only 32.3% of the PhoH profile matches AAK49953.1 (LBDG13). Irrespective of whether this PhoH match is a true or false positive, PhoH has no connection to the Nuclear Hormone receptor Ligand Binding Domain family and thus CDD does not annotate AAK49953.1 (LBDG13) as containing a Nuclear Hormone Receptor Ligand Binding Domain using CDD.

NCBI provides a public domain PSI-Blast server. Querying NCBI PSI-Blast with AAK49953.1 (LBDG13) through 10 positive iterations fails to annotate any region of AAK49953.1 (LBDG13) as having a relationship to any known Nuclear Hormone Receptor Ligand Binding Domains (note that NCBI PSI-Blast cannot provide data on negative iterations because no all-by-all calculation is performed). Figure 32 shows the graphical display of NCBI PSI-Blast results for AAK49953.1 (LBDG13). Figure 33 lists these NCBI PSI-BLAST hits (marked by lines). None of these sequences have been annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus NCBI PSI-Blast does not annotate AAK49953.1 (LBDG13) as having a relationship to any known Nuclear Hormone receptor Ligand Binding Domains.

The National Centre for Biotechnology Information (NCBI) GenBank protein database is viewed to examine if there is any further information that is known in the public domain relating to AAK49953.1 (LBDG13). This is the U.S. public domain database for protein and gene sequence deposition (Figure 33). AAK49953.1 was cloned by a group of scientists at the Indiana University School of Medicine, USA. The authors annotate AAK49953.1 (LBDG13) as a protein that is similar to CAC17441. CAC17441 is annotated as being similar to p65 protein, but as discussed earlier, p65 (CAA05410.2 (LBDG5)) is not annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus the (NCBI) GenBank does not in any way annotate AAK49953.1 (LBDG13) as containing a Nuclear Hormone Receptor Ligand Binding Domain or suggest that the protein has this function.

There is no further public domain annotation for AAK49953.1 (LBDG13). The public domain information for this protein does not annotate it as containing a Nuclear Hormone

Receptor Ligand Binding Domain. Therefore using all public domain annotation tools, AAK49953.1 (LBDG13) is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain.

To summarise we annotate residues 117-317 of AAK49953.1 (LBDG13) as adopting the structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 vn(LBDG5) which Genome Threader and Inpharmatica PSI-BLAST predict will adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain.